Genomic Imprinting in *Arabidopsis thaliana* – How Is It Regulated and Where Does It Occur?

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von

Michael Thomas Raissig
von
Bolligen BE und Meilen ZH

Promotionskomitee

Prof. Dr. Ueli Grossniklaus

Prof. Dr. Beat Keller

Prof. Dr. Wolf Reik

Zürich, 2013

TO MY FAMILY

ACKNOWLEDGMENTS

First of all I would like to thank **Ueli Grossniklaus** for being my scientific mentor with an open ear for scientific, professional or private problems. I am grateful for the given opportunity to do a PhD in his department and to work in such a friendly and stimulating environment.

I want to thank **Prof. Dr. Beat Keller** (University of Zürich) and **Prof. Dr. Wolf Reik** (Babraham Institute, Cambridge, UK) for being part of my thesis committee, reading and evaluating my thesis and for being co-examiners during my defense.

Then, I would like to thank **Célia Baroux**. She was my supervisor during my master thesis and from her I learned basically all "daily-business" lab skills. In addition, she introduced me to scientific thinking, study design and scientific debate.

I especially thank **Heike Lindner** for her love and her support, which was essential during the time of FUKUSHIMA in 2011 and throughout my thesis.

I thank the MEA4(N)EVA club (Presidente **Valeria**, Vice **Célia**, **Marian**, **Nuno**, **Heike W.**, **Wenjing**, **Milka**, **Johan**, **Anja**, **Alex**, **Olivier** and many master students) for scientific input, discussions, support and regular "Gipfeli".

For technical support I want to thank **Valeria** for running the lab (grazie!), **Daniela** for sequencing, **Arturo** for ordering, for music and for being my personal plant assistant and **Peter** for solutions, media and cleaning. And I would like to thank **Stefano** for cheering me up on a daily basis.

For friendship, chats and (lots of) parties, I would like to thank **Heike** (again!), **Deborah**, **Sibylle**, **Lena**, **Daniela**, **Valeria**, **Sharon**, **Milka**, **Anna**, **Rita**, **Célia**, **Wenjing**, **Evelyne**, **Heike W.**, **Stefan** a.k.a. Verstöpfeli, **Kostas** a.k.a. Costini, **Mayank** a.k.a. Puru, **Sailer** a.k.a. Ösi, **Arco** a.k.a. Don Dicko, **Johan** a.k.a. deejay, **Arturo** a.k.a. Burro a.k.a Penner, **Mo** (thanks for the screen), **Martin Bayer**, **Hampe**, **Aurel**, **Afif**, **Heichi**, **Nuno**, **Hannes**, **Sämi**, **Dyma**, **Marc**, **Manu**, **Marek**, **Christof**, and **Bruno** (thanks for keeping me motivated beginning 2010 drinking whiskey in my kitchen at 2am!).

Last, I want to thank my **Mum**, Jetty, my **Dad**, Peter, my stepmum, **Maria**, and my **brother**, Christian. I would not be who I am and where I am at the moment without them, thanks for everything.

ABSTRACT

Most genes are present in two copies in diploid cells, one of maternal and one of paternal origin. Usually, both copies share the same fate regarding their activity with both being either active or silent. Exceptions to this rule are genes regulated by genomic imprinting, where one gene copy is expressed and the other remains silent depending on the parent it was inherited from. The two copies are equal in their DNA sequence but carry distinct epigenetic modifications, an "imprint", and can therefore be distinguished. Genomic imprinting has independently evolved in mammals and in flowering plants, both of which share the placental habit, where the offspring develops within the mother and depends on her resources.

In mammals, many imprinted genes are clustered in groups and controlled by an imprinting control region (ICR) that is differentially marked according to its parental origin. The first identified plant ICR, the *MEDEA*-ICR, functions differently from mammalian ICRs. To understand how genomic imprinting is regulated at the *MEDEA* locus, we executed genetic and biochemical screens to identify novel imprinting regulators. Genetically, we identified two lethal mutations affecting *MEDEA* reporter gene expression.

Moreover, genomic imprinting in plants was thought to be restricted to the endosperm, a nutritive tissue similar in function to the mammalian placenta. Although genomic imprinting is most prominent in the placenta in mammals, it exists as well in the embryo and other tissues. By analyzing hybrid embryos, we could describe genes regulated by genomic imprinting in the embryo of the model plant *Arabidopsis thaliana*. Furthermore, we found that an epigenetic silencing complex, *Polycomb* Repressive Complex 2 (PRC2), is involved in maintaining the imprinted expression pattern at some loci. However, imprinted expression in the embryo requires erasure and resetting of the imprinting marks between the generations. Interestingly, the embryonically imprinted genes are expressed from both alleles in the seedling, suggesting that the imprint is erased during late embryogenesis or early seedling development.

The aim of this thesis was to elucidate regulation and lineage-specificity of genomic imprinting in *Arabidopsis.* We uncovered novel mutants involved in *MEDEA* reporter gene expression and seed viability, and confirmed the presence of genomic imprinting in plant embryos.

ZUSAMMENFASSUNG

Fast jedes Gen existiert in zwei Kopien in diploiden Zellen, einer mütterlichen und einer väterlichen, und normalerweise sind beide Kopien entweder aktiv oder inaktiv. Eine Ausnahme bilden Gene, die durch genomische Prägung reguliert sind, wobei eine Kopie aktiv und die andere inaktiv ist, und zwar je nachdem von welchem Elternteil die Genkopie geerbt wurde. Die Gene unterscheiden sich nicht in ihrer Sequenz, sind aber „geprägt“ durch epigenetische Markierungen. Genomische Prägung entstand sowohl in Säugern als auch in Pflanzen. Beide Gruppen haben eine ähnliche Fortpflanzungstrategie, wobei sich die Nachkommen innerhalb der Mutter und aufgrund ihrer Ressourcen entwickeln.

In Säugern befinden sich geprägte Gene in Gruppen und werden durch ein Steuerelement kontrolliert, das auf dem mütterlichen und dem väterlichen Chromosom unterschiedlich markiert ist. Das einzig bis anhin identifizierte Steuerelement in Pflanzen scheint anders zu funktionieren. Wir haben durch genetische Experimente Mutanten identifiziert, die einen Reporter eines geprägten Genes nicht richtig aktivieren und keine lebensfähigen Samen bilden können.

Obwohl in Säugern genomische Prägung in der Plazenta, im Embryo und in anderen Geweben vorkommt, dachte man in Pflanzen bis jetzt, dass genomische Prägung nur im Endosperm, der pflanzlichen „Plazenta“, existiert. Wir identifizierten „geprägte“ Gene auch in Embryonen der Modellpflanze *Arabidopsis thaliana*. Einige der „geprägten“ Gene im Embryo werden durch einen repressiven epigenetischen Komplex (PRC2) reguliert. Interessanterweise, sind die Gene im *Arabidopsis* Keimling nicht mehr geprägt, was darauf schliessen lässt, dass die „Prägung“ spät in der Embryonalentwicklung oder früh in der Keimlingsentwicklung gelöscht wird.

Das Ziel dieser Dissertation war, mehr über das „wie“ und das „wo“ von genomischer Prägung in Pflanzen herauszufinden. Zum einen fanden wir Mutationen, die die Regulierung von geprägten Genen und die Samenentwicklung beeinflussen, und zum anderen zeigten wir, dass genomische Prägung auch im Pflanzenembryo vorkommt.

TABLE OF CONTENTS

INTRODUCTION

SCOPE OF THE THESIS

Genomic Imprinting in *Arabidopsis thaliana* - How & Where?

Genomic imprinting is an epigenetic phenomenon and leads to monoallelic gene expression that is dependent on its parent-of-origin. Usually, the maternal and the paternal allele of a given gene share the same fate in a specific tissue, meaning that they are either both expressed or both silent. Important exceptions to this rule are genes regulated by genomic imprinting, where only one, or preferentially one, allele is expressed and the other remains silent or weakly expressed. The two alleles do not differ in their sequence but rather carry parent-specific, epigenetic imprints that allow the cell to distinguish the two alleles even if they are in the same nucleus (Reik and Walter, 2001; Bartolomei and Ferguson-Smith, 2011; Ferguson-Smith, 2011; Barlow, 2011; Raissig et al., 2011). Genomic imprinting independently evolved in mammals and flowering plants (angiosperms). In mammals, genomic imprinting occurs mainly in the placenta, but also in the embryo, and even in adult tissues like the brain (Reik and Walter, 2001; Bartolomei and Ferguson-Smith, 2011; Ferguson-Smith, 2011; Barlow, 2011). Yet, in flowering plants, genes regulated by genomic imprinting seem to be largely restricted to the endosperm, a nutritive tissue similar to the mammalian placenta (Raissig et al., 2011). Regulation of genomic imprinting in mammals relies on imprinting control regions (ICRs) that are differentially marked in the germ line and control whole gene clusters through epigenetic processes including DNA methylation, posttranslational histone modification and specific, monoallelic expression of long non-coding RNAs (ncRNAs, see below; Barlow, 2011; Bartolomei and Ferguson-Smith, 2011; Ferguson-Smith, 2011). In angiosperms, imprinted genes are not organized in gene clusters but are rather singletons, and allele-specific expression is regulated by DNA methylation and the *Polycomb* Repressive Complex 2 (PRC2; see later, Raissig et al. 2011). Thus far, only one plant ICR has been identified, the *MEDEA (MEA)*-ICR, that functions independently of DNA methylation (see later; Wöhrmann et al., 2012).

The main research questions addressed in this thesis are how and where genomic imprinting occurs in the model organism *Arabidopsis thaliana*:

- How is genomic imprinting of *MEA* controlled at the *MEA*-ICR, which functions independently of DNA methylation?
- What are the *trans*-acting factors activating or repressing *MEA* expression by allele-specific binding to the *MEA*-ICR?
- Is genomic imprinting really restricted to the endosperm in *Arabidopsis*?
- Do we find genes regulated by genomic imprinting also in the *Arabidopsis* embryo?

Part 1 of this thesis describes the identification of the *MEA*-ICR by a former PhD student (Wöhrmann et al., 2012; Chapter 1.1), the attempts to find *trans*-acting regulators controlling the *MEA*-ICR (Chapter 1.2) and SNP ratio mapping (SRM), a method that we developed to map gametophytic or homozygous-lethal mutations by next-generation sequencing (Lindner et al., 2012; Chapter 1.3). Part 2 first presents a novel method facilitating the isolation of early-stage embryos from *Arabidopsis* seeds and allowing the study of embryonic tissue in detail (Raissig et al., in press, Chapter 2.1). Second, we present the assessment of parental contribution to the early embryonic transcriptome and its regulation (Autran et al., 2011; Chapter 2.2) that suggested the presence of monoallelic, parent-of-origin-dependent gene expression in the *Arabidopsis* embryo. In addition, we discuss a recent report (Nodine and Bartel, 2012) that challenged the maternal dominance in the early embryonic transcriptome described by our group. Lastly, we present the first confirmation of genes regulated by genomic imprinting in the *Arabidopsis* embryo and show that this epigenetic phenomenon is not a unique feature of the endosperm (Raissig

et al., in preparation; Chapter 2.3).

Throughout this thesis published work is inserted as layouted by the journals (if no copyright issues exist and/or permissions to reprint articles have been granted) and is preceded by a short note indicating my contribution to the work. Supplemental information to all publications can be found in the Appendix. Every chapter – even if it is not published (yet) - is written in the form of a publication including introduction, material and methods, results and discussion and is followed by the referred literature for each chapter. Therefore, references might be cited more than once.

References

Autran, D., Baroux, C., Raissig, M.T., Lenormand, T., Wittig, M., Grob, S., Steimer, A., Barann, M., Klostermeier, U.C., Leblanc, O., Vielle-Calzada, J.-P., Rosenstiel, P., Grimanelli, D., and Grossniklaus, U. (2011). Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. Cell **145**: 707–719.

Barlow, D.P. (2011). Genomic imprinting: a Mammalian epigenetic discovery model. Annu Rev Genet **45**: 379–403.

Bartolomei, M.S. and Ferguson-Smith, A.C. (2011). Mammalian genomic imprinting. Cold Spring Harb Perspect Biol **3**.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet **12**: 565–575.

Lindner, H., Raissig, M., Sailer, C., Shimosato-Asano, H., Bruggmann, R., and Grossniklaus, U. (2012). SNP-Ratio Mapping (SRM): Identifying Lethal Alleles and Mutations in Complex Genetic Backgrounds by Next-Generation Sequencing. Genetics **191**: 1381–1386.

Nodine, M.D. and Bartel, D.P. (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature **482**: 94–97.

Raissig, M., Gagliardini, V., Jaenisch, J., Grossniklaus, U., and Baroux, C. Efficient and rapid isolation of early-stage embryos from *Arabidopsis thaliana* seeds. J Vis Exp *in press*

Raissig, M.T., Baroux, C., and Grossniklaus, U. (2011). Regulation and Flexibility of Genomic Imprinting during Seed Development. Plant Cell **23**: 16–26.

Reik, W. and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat Rev Genet **2**: 21–32.

Wöhrmann, H.J., Gagliardini, V., Raissig, M.T., Wehrle, W., Arand, J., Schmidt, A., Tierling, S., Page, D.R., Schöb, H., Walter, J., and Grossniklaus, U. (2012). Identification of a DNA methylation-independent imprinting control region at the *Arabidopsis MEDEA* locus. Genes Dev **26**: 1837–1850.

INTRODUCTION
CHAPTER 1

GENOMIC IMPRINTING IN MAMMALS

A historical perspective on the discovery of genomic imprinting – from chromosomes to genes

Chromosomal imprinting was the first form of "imprinting" described in the literature. A zygote of the dipteran sciarid fly inherits one maternal X chromosome and two paternal X chromosomes. In female embryos, one of the two paternal X chromosomes is eliminated, whereas in males both paternal X chromosomes are lost. Specific removal of X chromosome(s) that are always of paternal origin suggested that the paternal chromosomes must retain a memory of their parental origin, a so-called "imprint". Crouse (1960) realized that a heterochromatic region on the X chromosome is essential for "marking" the parental origin of the chromosome and for the specific removal of the paternal chromosome(s) (Crouse, 1960). Similarly, in mealy bugs, the entire paternal set of chromosomes is compacted into heterochromatin and remains silent throughout development (Brown and Nur, 1964). In mammals, dosage-compensation of X-linked genes between males (one X) and females (two X) is regulated by X chromosome inactivation (XCI), a form of chromosomal imprinting (reviewed in Lee, 2011). XCI is either imprinted or random regarding the silencing of the maternal or the paternal X chromosome, depending on the organism or the tissue. Imprinted X chromosome inactivation occurs in female marsupials (Cooper et al., 1971; Sharman, 1971) and in the extra-embryonic lineage of mice, where always the paternally inherited X-chromosome gets inactivated (Takagi and Sasaki, 1975). The X-inactivation center (Xic) initiates inactivation by producing the non-coding *Xist* transcript, which silences the chromosome in *cis* by recruiting *Polycomb* Repressive Complex2 (PRC2), an epigenetic silencer, and by coating the inactive chromosome (Lee, 2011).

In 1970, reciprocal crosses of maize varieties with differently colored kernels led to the discovery that also single gene activity can depend on the parent-of-origin. If the maize *R1* gene, which regulates anthocyanin synthesis in the seed, is inherited maternally the kernels are fully colored, whereas only a mottled coloring is observed when the *R1* gene is inherited paternally (Kermicle, 1970). After the pioneering work of the maize geneticist Kermicle (1970), a series of experimental evidence in the 1980s led to the assumption that the parental genomes are not equivalent in mouse: First, gynogenotes and androgenotes created by nuclear transfer experiments in mouse oocytes, where either two maternal or two paternal pronuclei were transferred to enucleated egg cells, respectively, failed to complete development (Barton et al., 1984; McGrath & Solter, 1984; Surani et al., 1984). Then, Cattanach and Kirk (1985) found that some uniparental disomies in mouse develop normally, whereas others abort embryo development or have other parent-of-origin-specific defects (Cattanach and Kirk, 1985). Therefore, it was concluded that not the entire genome is subject to parental effects but that rather single chromosomes or chromosomal regions harbor imprinted loci. In fact, the assumption that absence or overexpression of imprinted genes underlies the developmental failure of andro- and gynogenotes was largely shown to be true by combining nuclei of two different oocytes that had mutations in two different imprinted loci. Such bimaternal offspring had a normal imprinted gene dosage and was viable (Kono et al., 2004; Kawahara et al., 2007).

Igf2r (insulin-like growth factor type 2 receptor gene) was the first endogenous imprinted gene identified in mammals, in the scope of determining which gene was responsible for the lethality of mice that maternally inherited a deleted segment of chromosome 17, the Hairpin-tail deletion (Barlow et al., 1991; Johnson, 1974). Thereafter, the non-coding mouse gene *H19* was shown to be imprinted by tracking the polymorphic transcript in hybrid mice (Bartolomei et al., 1991), and *Igf2* (Insulin-like growth factor 2) was found to be imprinted in a gene targeting approach, where mice inheriting the deleted gene paternally were phenotypically like homozygous null mice for the *Igf2* gene, whereas the maternally inherited deletion had no effect at all (DeChiara et al., 1991). Imprinted expression of *Igf2* was confirmed by showing absence of the *Igf2* gene product in mouse embryos with a maternal uniparental disomy of distal chromosome 7, suggesting paternal expression only (Ferguson-Smith et al., 1991). Nowadays, approximately 100 imprinted genes have been identified in mammals, most of which reside in gene clusters, are regulated within the cluster, and mostly contain a non-coding RNA (ncRNA) and both maternally expressed and paternally expressed genes (MEGs and PEGs). For a complete list, see http://igc.otago.

ac.nz/home.html or http://www.mousebook.org/.

Evolution of genomic imprinting – a parental tug-of-war

Genomic imprinting evolved independently in mammals and in flowering plants (angiosperms). Both organisms share a common reproductive strategy, the placental habit, where the embryo develops within the reproductive apparatus of the mother and depends on her nutrients (Figure I1). Maternal support of embryo development depends on the formation of an extra-embryonic, nutrient-transferring tissue – the placenta in mammals and the endosperm in plants. Subsequently, parental investment towards embryogenesis is extremely unbalanced and, in fact, solely depends on the mother. Therefore, an intragenomic parental conflict over resource allocation to the developing offspring was proposed to explain the evolution of genomic imprinting in both, mammals and flowering plants (Haig and Westoby, 1989; Moore and Haig, 1991). It predicts that genes restricting growth have evolved to be under maternal control to guarantee equal distribution of maternal resources to all present and future offspring. In contrast, the paternal genes intend to maximize nutrient transfer and embryonic growth to enhance offspring fitness, since it is not at the father's expense and because he sires only a subset of all offspring of a polyandric mother. Although several other theories were formulated to explain the evolution of genomic imprinting, there is quite some evidence supporting the intragenomic parental conflict theory: First, many imprinted genes in both, mammals and plants, are predominantly expressed in the mammalian placenta and the plant endosperm (Frost & Moore, 2010; Jullien & Berger, 2009; Raissig et al., 2011; Reik & Walter, 2001).

Second, functional studies on knock-out mice and mutant plants revealed a role of imprinted genes in regulating growth of the embryo and the extra-embryonic placenta or endosperm in a parent-of-origin-dependent manner. The reciprocally imprinted mammalian genes *Igf2* and *Igf2r* provide a remarkable example illustrating the parental conflict during early development: While mutant mice with a disrupted *Igf2* gene, which is normally paternally expressed, show placental and fetal growth restriction phenotypes, knock-out mice with a disrupted maternal *Igf2r*, display the opposite phenotype (Lau et al., 1994; Ludwig et al., 1996; Constância et al., 2002; Reik et al., 2003; Tycko and Morison, 2002; Angiolini et al., 2006; Frost and Moore, 2010). Remarkably, specific knock-out of placental *Igf2* (P0) results in an underdeveloped placental but normal fetal development almost until term (Constância et al., 2002), suggesting that IGF-2 is involved in regulating two aspects of fetal growth, namely the supply potential (by acting in the placenta) and the supply demand (by acting in the fetus). The tug-of-war between the parental genomes in supplying fetal growth is further illustrated by knock-out mice in several additional imprinted genes, including the mutants in the maternally expressed genes *Phlda2*, *Grb10* and *Cdkn1c* of the *Kcnq1* imprinting cluster that all display placental overgrowth and sometimes fetal overgrowth phenotypes (Angiolini et al., 2006, Frost & Moore, 2010). In plants, the *MEDEA* (*MEA*) gene is maternally expressed in the endosperm and, if disrupted, leads to embryo overgrowth and, therefore, fits the prediction that maternally expressed genes are involved in restricting growth also in plants (Grossniklaus et al., 1998).

Third, no evidence exists for imprinting in egg-laying or oviparous mammals, the monotremes, like the platypus or the echidna (Suzuki et al., 2007; Edwards et al., 2008), but some of the imprinted domains like the *Igf2-H19* and the *Igf2r* clusters are imprinted in marsupials, already having a yolk sac and/or a placenta (Renfree et al., 2009; Smits et al., 2008). In addition, imprinted XCI is restricted to marsupials and to the placenta of eutherian mammals (Cooper et al., 1971; Sharman, 1971; Takagi and Sasaki, 1975). Lastly, imprinting is generally thought to be absent from egg-laying animal classes like reptiles, amphibians or fish (Reik et al., 2003).

Other driving forces and theories were proposed to explain the evolution of genomic imprinting and some of them are supported by experimental evidence and, therefore, require to be mentioned: The meiotic recombination theory proposes a mechanistic link between meiotic pairing of chromosomes and genomic

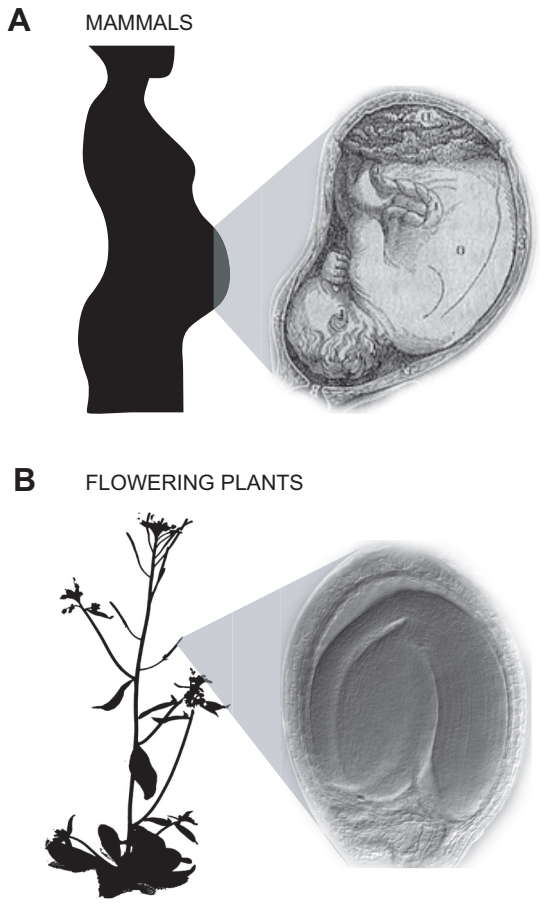


Figure I1. "Placental habit" as the common mode of reproduction between mammals and flowering plants. Placental reproduction or the "placental habit" is an outstanding example of convergent evolution between mammals (A) and flowering plants (B). The next generation develops within the mother and is solely dependent on maternal resources provided through the mammalian placenta or the plant endosperm, whereas the father has no cost during embryogenesis. Thus the mother has to restrict embryo growth to equally distribute her resources to all her (present and future) offspring. The father, however, has an interest in promoting growth of the embryo as he has no costs, and is likely not genetically related to all offspring of a polyandric mother. As a consequence, maternally expressed genes have evolved to restrict growth (like *Igf2r* or *MEDEA*), whereas paternally expressed genes have evolved to promote growth (like *Igf2*). This evolutionary tug-of-war is called the parental conflict or kinship theory. Source of (modified) human embryo drawing: Wikimedia Commons (http://commons.wikimedia.org).

imprinting. Reciprocally imprinted domains might aid in the recognition process between homologous chromosomes and could prevent pairing with non-homologous chromosomes (de Villena et al., 2000). This theory is supported empirically by the observation that imprinted loci of human chromosomes show sex-specific recombination frequencies (Pàldi et al, 1995; Robinson & Lalande, 1995), and that imprinted regions in the human genome are recombination hotspots (Lercher and Hurst, 2003; Sandovici et al., 2006). In contrast, imprinting might have evolved as a side-effect of silencing foreign DNA elements, namely transposable elements (TEs), since mammalian ICRs share DNA-sequence features that are normally attributed to TEs and, in addition, many TEs can be found up- or downstream of plant imprinted genes (Gehring et al., 2009; Kinoshita et al., 2004; Wolff et al., 2011). Thus, regulation of imprinted expression could be considered as an "extension of host defense" (Barlow, 1993). Other theories like the ovarian time bomb theory, which suggests that imprinting evolved to prevent parthenogenesis (Varmuza and Mann, 1994), and the maternal-offspring co-adaptation theory (Wolf and Hager, 2006), predicting that genes acting at the interface between mother and offspring have evolved to be under maternal control, are discussed in other chapters of this thesis.

In summary, the parental-conflict theory presents the best and most widely accepted theory for the convergent evolution of imprinting in flowering plants and mammals, since other theories fail to explain certain inconsistencies. For example, some TEs in the vicinity of plant imprinted genes are dispensable for imprinting, like the helitron upstream and the differentially methylated direct repeats downstream of *MEA* (Spillane et al., 2004). And, although there is evidence for the meiotic recombination theory, this is generally applicable to all sexual organisms and not restricted to plants and mammals, and thus not explaining the specific emergence of imprinting in those two groups only.

Biological function of imprinted genes in mammals – growth, development, and behavior

Imprinted genes in mammals have been implicated in placental and embryonic growth control, in the development of particular lineages, in normal brain function, and even in postnatal effects, such as energy homeostasis and behavior.

As mentioned before, many imprinted genes are expressed in the placenta and/or the embryo and are

directly involved in regulating nutrient transfer to the developing embryo by modulating placental size, transport activity, or nutrient demand in the embryo (Tycko and Morison, 2002; Angiolini et al., 2006; Reik et al., 2003; Frost and Moore, 2010). Apart from the severe effects of misexpressed *Igf2* and *Igf2r* on placental and fetal growth described above, many other imprinted genes influence prenatal growth or placental differentiation like several maternally expressed genes - imprinted in the trophoblast only – of the imprinted gene cluster *Kcnq1* on mouse chromosome 7. For example, a mutation in maternally expressed *Ascl2* (aka. *Mash2*), which is implicated in the differentiation of the placenta, is embryonic lethal (Guillemot et al., 1995). In addition, maternally expressed *Cdkn1c* is a growth repressor and mutations in this gene lead to neonatal lethality in the mouse and can cause the Beckwith-Wiedemann Syndrome (BWS) in humans (Choufani et al., 2010; Yan et al., 1997; see below). *Phlda2* is another maternally expressed growth inhibitor and abolition of its expression leads to placental hyperplasia, whereas doubling its dose by disrupting imprinted expression results in placental and fetal growth retardation (Frank et al., 2002; Tunster et al., 2010). Just recently, the role of *Phlda2* in growth control was directly linked to glycogen storage and nutrient transfer from placenta to fetus (Tunster et al., 2010).

Although genomic imprinting is particularly prevalent during prenatal growth and development of placenta and fetus, it has also a strong implication after birth. In the brain, the imprinted gene $G_s\alpha$ is maternally expressed in the hypothalamus and controls energy expenditure. Mice with a brain-specific deletion of the maternally inherited allele of $G_s\alpha$ are glucose-intolerant, insulin-resistant and obese (Chen et al., 2009). Even more striking, mothers that are mutant in two strongly paternally expressed genes in the brain, *Peg1* and *Peg3*, show reduced maternal care, suggesting postnatal grand-paternal influence on maternal offspring care (Lefebvre et al., 1998; Li et al., 1999). In addition, the expression of many imprinted transcripts in the brain is consistent with the neurodevelopmental effects of some human disorders, like the behavioral and neurodevelopmental disorders Prader-Willi Syndrome (PWS) and Angelman Syndrome (AS). Both complex but distinct disease syndromes are associated with the human chromosomal region 15q11q13 (Buiting, 2010). While AS can be caused by a mutation in the maternally expressed gene *UBE3A* in the brain, the situation is less clear for PWS (Buiting, 2010). Yet, a deficiency concerning the paternally expressed *SNORD116* small nucleolar RNA (snoRNA) gene can cause PWS, but other genes might add to the syndrome (Buiting, 2010). However, both syndromes are mostly caused by uniparental disomies or chromosomal deletions in the 15q11q13 region (Bartolomei & Ferguson-Smith, 2011; Buiting, 2010).

In contrast to AS and PWS, imprinted genes affecting placental or fetal growth underlie the imprinted growth disorders BWS and Silver-Russell Syndrome (SRS; Bartolomei & Ferguson-Smith, 2011; Choufani et al., 2010; Eggermann, 2010). BWS, an overgrowth disorder including an increased risk of childhood tumors can be caused by a mutation in cyclin-dependent kinase inhibitor 1C (*CDKN1C*), a maternally expressed imprinted cell cycle regulator. Alternatively, epimutations like hypermethylation of the maternally unmethylated ICR or a paternal uniparental disomy for chromosome 11 can underlie BWS, leading to biallelic expression and, therefore, to a double dose of *IGF2* (Choufani et al., 2010). In contrast, the main phenotype of SRS is intra-uterine growth restriction and is often associated with hypomethylation of the paternal ICR, which results in a reduction of *IGF2* transcription (Eggermann, 2010).

Taken together, genomic imprinting plays an important role in growth, development, and behavior throughout the life cycle of mammals, although there are probably not many more than a 100 imprinted genes in the mammalian genome. Apparently, balancing fetal growth and postnatal behavior via parent-of-origin-dependent monoallelic gene expression is important enough to risk potentially lethal or harmful consequences due to haploinsufficiency.

Regulation of genomic imprinting in mammals – a tale of gene clusters, ICRs, and ncRNAs

Imprinted genes in mammals usually reside in gene clusters that are regulated by a single imprinting control region (ICR), which was identified by targeted deletion of differentially methylated regions (DMRs; Lin

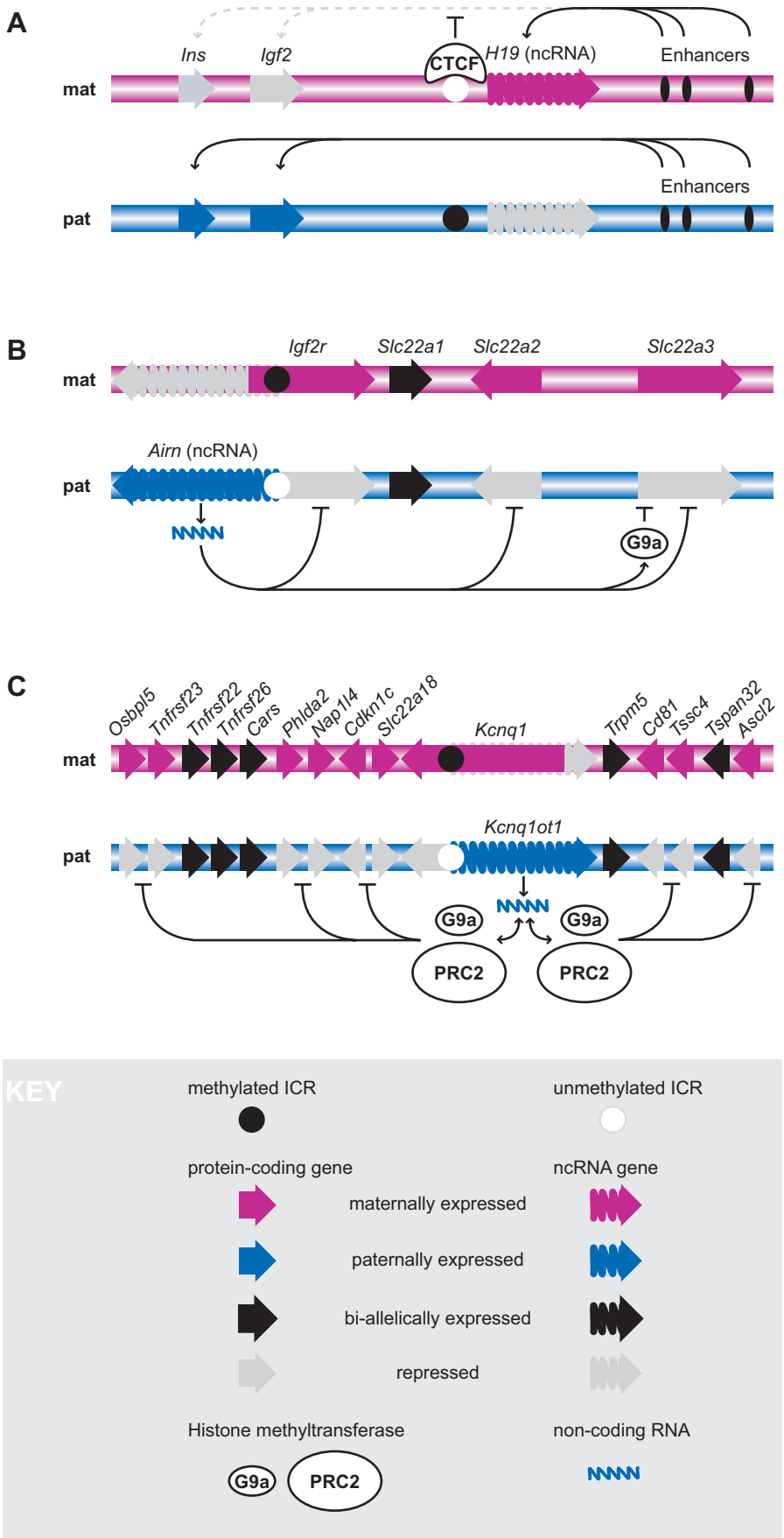


Figure I2. Regulation of imprinted gene clusters in mammals. Differential DNA methylation of ICRs, either paternal methylation (A) or maternal methylation (B and C), regulate imprinted gene clusters by different means. (A) *Igf2-H19* imprinted gene cluster: CTCF, a zinc-finger insulator protein, binds only to the unmethylated maternal ICR, whereby it prevents the interaction of enhancers with the two protein-coding genes *Igf2* and *Ins* in *cis*. However, the enhancer interacts with the maternal *H19* allele, activating its expression. On the paternal chromosome, binding of CTCF to the ICR is prevented by DNA methylation, and the enhancers are able to interact with and activate the paternally expressed genes *Igf2* and *Ins*. (B) *Igf2r* imprinted gene cluster: The unmethylated paternal ICR acts as a promoter to express the ncRNA *Airn*, which itself represses the paternal alleles in *cis* by antisense transcription, interaction with chromatin, and recruiting the H3K9 histone methyltransferase G9a to *Slc22a3*. On the maternal chromosome the ICR is methylated and cannot drive expression of *Airn*. Hence, the maternal alleles of *Igf2r*, *Slc22a2* and *Slc22a3* are expressed. (C) *Kcnq1* imprinted gene cluster: The unmethylated paternal ICR acts as a promoter to express the ncRNA *Kcnq1ot1*, which itself represses the paternal alleles in *cis* by interaction with chromatin and recruiting G9a and PRC2. On the maternal chromosome, the methylated ICR prevents expression of the ncRNA *Kcnq1ot1* and the maternal alleles are expressed. The clusters are not drawn to scale. mat: maternal chromosome; pat: paternal chromosome.

et al., 2003; Smilinich et al., 1999; Sutcliffe et al., 1994; Thorvaldsen et al., 1998; Williamson et al., 2006; Wutz et al., 1997). ICRs are marked by DNA methylation in a sex-specific manner in the primordial germ cells (PGCs), where the chromosomes are in distinct and separated compartments. During PGC development, the whole genome undergoes demethylation that is followed by a sex-specific remethylation of the ICRs by the *de novo* DNA methyltransferase DNMT3A and its cofactor DNMT3L (Bourc'his et al., 2001; Kaneda et al., 2004; Reik, 2007). How exactly the epigenetic machinery recognizes ICRs is still unclear. Either specific spacing of CpG islands within the ICR, specific histone modifications at ICRs or parent-specific transcription over ICRs could guide DNA methyltransferases to the ICR (Chotalia et al., 2009; Ciccone et al., 2009; Glass et al., 2009; Ooi et al., 2007). After setting the imprints in the germline, they have to be maintained after fertilization and during the subsequent embryonic divisions. Especially in the zygote, when the gametes' genomes are rapidly and extensively reprogrammed and DNA methylation and chromatin modifications are erased and reset to acquire a pluripotent state, the imprints must escape resetting (Morgan et al., 2005). Paternally inherited chromatin is actively demethylated, likely via hydroxymethylcytosine intermediates (Iqbal et al., 2011; Wossidlo et al., 2011), followed by passive demethylation of both parental genomes during cleavage and replication in the absence of DNMT1o, the oocyte-specific form of the maintainance DNA methyltransferase DNMT1. DNMT1o is excluded from the nucleus except at the 8-cell embryo stage, where nuclear localization of DNMT1o might be responsible for maintaining the imprints (Howell et al., 2001). However, mutant studies implicated also the somatic form of DNMT1, DNMT1s, in maintaining imprints during embryogenesis (Li et al., 1993). Furthermore, additional *trans*-acting factors like DPPA3 or ZFP57 seem to be involved in the maintenance and/or protection of the epigenetic modification(s) at ICRs during global reprogramming (Li et al., 2008; Nakamura et al., 2007). DPPA3 (or Stella) is a DNA-binding protein and is highly abundant in oocytes and enters both pronuclei after fertilization (Nakamura et al., 2007). Maternal deletion of the gene causes embryonic lethality and loss of DNA methylation at some imprinted genes, suggesting that DPPA3 specifically protects DMRs or even ICRs from demethylation (Nakamura et al., 2007). Moreover, a mutation in the KRAB zinc finger gene *Zfp57*, a transcriptional repressor, is lethal and causes aberrant DNA methylation and expression of a number of imprinted loci (Li et al., 2008). Interestingly, zygotic expression of *Zfp57* could partially rescue the lack of DNA methylation at the *Snrpn* ICR, which is usually set in the oocyte and henceforth protected by maternal ZFP57, and reestablishes proper DNA methylation patterns during embryonic development (Li et al., 2008). This indicates that the *Snrpn* ICR harbored a primary imprinting mark, which is distinct from DNA methylation and is sufficient to direct *de novo* DNA methylation at the maternal *Snrpn* ICR after fertilization (Li et al., 2008). Generally, it could be hypothesized that unique epigenetic modifications distinct from DNA methylation might guide the gametic establishment and the embryonic maintenance of differentially methylated ICRs during reproductive development.

The differentially methylated ICRs are key to mammalian imprinting regulation. ICRs can be paternally or maternally methylated and either function as insulator elements, regulate enhancer-promoter interactions, or act as promoters of ncRNAs, which themselves can repress protein-coding genes in *cis* (Figure I2). The ICR of the *H19-Igf2* locus functions as an insulator element and regulates the interaction of DNA regulatory elements in *cis*. A zinc-finger insulator protein, CTCF, binds exclusively to the unmethylated maternal ICR (Bell & Felsenfeld, 2000; Hark et al., 2000; Szabó et al., 2000). Binding of CTCF prevents the interaction of an enhancer with the maternal *Igf2* allele, therefore preventing its expression but allowing, through enhancer interaction, maternal expression of the ncRNA *H19* (Engel et al., 2004; Murrell et al., 2004). On the paternal chromosome, CTCF cannot bind to the methylated ICR, which allows the interaction of the enhancer with the paternal *Igf2* allele to drive its expression (Figure I2A). In contrast, the maternally methylated ICRs of the *Igf2r* cluster and the *Kcnq1* cluster regulate imprinting by preventing expression of a repressive ncRNA from the maternal chromosome (Figure I2B and I2C). At the *Igf2r* cluster, the unmethylated ICR on the paternal chromosome acts as a promoter

for a long ncRNA, *Airn*, which is transcribed in antisense direction to the *Igf2r* gene (Wutz et al., 1997; Lyle et al., 2000). The antisense transcription overlaps the promoter of *Igf2r*, which is required for its paternal repression in *cis* (Sleutels et al., 2002), but two genes downstream of *Igf2r* are also repressed, although the antisense transcription of *Airn* is in the opposite direction of those two genes (Figure I2B). Recent evidence suggests that the ncRNA interacts in *cis* with the promoter of one of those genes and recruits a repressive histone modification machinery specifically to the paternal chromosome (Figure I2B; Nagano et al., 2008). At the *Kcnq1* cluster, the paternally unmethylated ICR promotes expression of the ncRNA *Kcnq1ot1*, which represses 11 genes in *cis* (Mancini-Dinardo et al., 2006). *Kcnq1ot1* interacts with chromatin and recruits the Histone 3 Lysine 9 (H3K9)- and Histone 3 Lysine 27 (H3K27)-specific histone methyltransferase activities G9a and PRC2, respectively (Figure I2C). Silencing of the 11 paternal genes in *cis* is correlated with enriched H3K9me3 and H3K27me3 levels at the *Kcnq1* cluster in the placenta (Lewis et al., 2004; Pandey et al., 2008; Redrup et al., 2009; Terranova et al., 2008; Umlauf et al., 2004).

Remarkably, some imprinted gene clusters are dynamically controlled in the course of development. The imprinted gene *Dlk1* in mice is exclusively expressed paternally during embryogenesis but is biallelically expressed in the neurogenic niche shortly after birth, therefore dramatically increasing its dosage (Ferrón et al., 2011). Both parental alleles are required for the maintenance of the neural stem cell pool and, thus, the generation of new neurons. The lineage-specific loss of imprinting correlates with the methylation of the usually unmethylated maternal ICR at this locus (Ferron et al., 2011). This suggests that modulating gene dosage in different developmental contexts can be achieved by dynamic epigenetic regulation of imprinted genes.

Taken together, differentially methylated ICRs in imprinted gene clusters in mammals adopt multiple roles depending on their methylation status: unmethylated ICRs can act as (i) *cis* repressors of protein coding genes in the form of an insulator, and (ii) as *cis* activators of ncRNAs. The ncRNAs repress protein-coding genes in *cis*, either by physically coating the locus and preventing transcription, as it is the case for X chromosome inactivation (Lee, 2011), or by recruiting repressive epigenetic machineries, or a combination of both (Pandey et al., 2008, Redrup et al., 2009, Terranova et al., 2008).

References

Angiolini, E., Fowden, A., Coan, P., Sandovici, I., Smith, P., Dean, W., Burton, G., Tycko, B., Reik, W., Sibley, C., and Constância, M. (2006). Regulation of placental efficiency for nutrient transport by imprinted genes. Placenta **27 Suppl A**: 98–102.

Barlow, D.P. (1993). Methylation and imprinting: from host defense to gene regulation? Science **260**: 309–310.

Barlow, D.P., Stöger, R., Herrmann, B.G., Saito, K., and Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the *Tme* locus. Nature **349**: 84–87.

Bartolomei, M.S. and Ferguson-Smith, A.C. (2011). Mammalian genomic imprinting. Cold Spring Harb Perspect Biol **3**.

Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse *H19* gene. Nature **351**: 153–5.

Barton, S.C., Surani, M.A., and Norris, M.L. (1984). Role of paternal and maternal genomes in mouse development. Nature **311**: 374–376.

Bell, A.C. and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. Nature **405**: 482–485.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science **294**: 2536–2539.

Brown, S.W. and Nur, U. (1964). Heterochromatic Chromosomes in the Coccids. Science **145**: 130–136.

Buiting, K. (2010). Prader-Willi syndrome and Angelman syndrome. Am J Med Genet C Semin Med Genet **154C**: 365–376.

Cattanach, B.M. and Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. Nature **315**: 496–498.

Chen, M., Wang, J., Dickerson, K.E., Kelleher, J., Xie, T., Gupta, D., Lai, E.W., Pacak, K., Gavrilova, O., and Weinstein, L.S. (2009). Central nervous system imprinting of the G protein G(s)alpha and its role in metabolic regulation. Cell Metab **9**: 548–55.

Chotalia, M., Smallwood, S.A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., James, K., Dean, W., and Kelsey, G. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev **23**: 105–117.

Choufani, S., Shuman, C., and Weksberg, R. (2010). Beckwith-Wiedemann syndrome. Am J Med Genet C Semin Med Genet **154C**: 343–354.

Ciccone, D.N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., and Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature **461**: 415–418.

Constância, M., Hemberger, M., Hughes, J., Dean, W., Ferguson-Smith, A., Fundele, R., Stewart, F., Kelsey, G., Fowden, A., Sibley, C., and Reik, W. (2002). Placental-specific IGF-II is a major modulator of placental and fetal growth. Nature **417**: 945–948.

Cooper, D.W., VandeBerg, J.L., Sharman, G.B., and Poole, W.E. (1971). Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X inactivation. Nature New Biol **230**: 155–157.

Crouse, H.V. (1960). The Controlling Element in Sex Chromosome Behavior in *Sciara*. Genetics **45**: 1429–1443.

DeChiara, T.M., Robertson, E.J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. Cell **64**: 849–859.

Edwards, C.A., Mungall, A.J., Matthews, L., Ryder, E., Gray, D.J., Pask, A.J., Shaw, G., Graves, J.A.M., Rogers, J., consortium, S., Dunham, I., Renfree, M.B., and Ferguson-Smith, A.C. (2008). The evolution of the DLK1-DIO3 imprinted domain in mammals. Plos Biol **6**: e135.

Eggermann, T. (2010). Russell-Silver syndrome. Am J Med Genet C Semin Med Genet **154C**: 355–364.

Engel, N., West, A.G., Felsenfeld, G., and Bartolomei, M.S. (2004). Antagonism between DNA hypermethylation and enhancer-blocking activity at the *H19* DMD is uncovered by CpG mutations. Nat Genet **36**: 883–888.

Ferguson-Smith, A.C., Cattanach, B.M., Barton, S.C., Beechey, C.V., and Surani, M.A. (1991). Embryological and molecular investigations of parental imprinting on mouse chromosome 7. Nature **351**: 667–70.

Ferrón, S.R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., Morante-Redolat, J.M., Laborda, J., Guillemot, F., Bauer, S.R., Fariñas, I., and Ferguson-Smith, A.C. (2011). Postnatal loss of *Dlk1* imprinting in stem cells and niche astrocytes regulates neurogenesis. Nature **475**: 381–385.

Frank, D., Fortino, W., Clark, L., Musalo, R., Wang, W., Saxena, A., Li, C.-M., Reik, W., Ludwig, T., and Tycko, B. (2002). Placental overgrowth in mice lacking the imprinted gene *Ipl*. Proc Natl Acad Sci USA **99**: 7490–7495.

Frost, J.M. and Moore, G.E. (2010). The importance of imprinting in the human placenta. PLoS Genet **6**: e1001015.

Gehring, M., Bubb, K.L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science **324**: 1447–1451.

Glass, J.L., Fazzari, M.J., Ferguson-Smith, A.C., and Greally, J.M. (2009). CG dinucleotide periodicities recognized by the Dnmt3a-Dnmt3L complex are distinctive at retroelements and imprinted domains. Mamm Genome **20**: 633–643.

Grossniklaus, U., Vielle-Calzada, J.P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal control of embryogenesis by *MEDEA*, a *polycomb* group gene in *Arabidopsis*. Science **280**: 446–450.

Guillemot, F., Caspary, T., Tilghman, S.M., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Anderson, D.J., Joyner, A.L., Rossant, J., and Nagy, A. (1995). Genomic imprinting of *Mash2*, a mouse gene required for trophoblast development. Nat Genet **9**: 235–242.

Haig, D. and Westoby, M. (1989). Parent-Specific Gene Expression and the triploid Endosperm. American Naturalist **134**: 147–155.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. Nature **405**: 486–489.

Howell, C.Y., Bestor, T.H., Ding, F., Latham, K.E., Mertineit, C., Trasler, J.M., and Chaillet, J.R. (2001). Genomic imprinting disrupted by a maternal effect mutation in the *Dnmt1* gene. Cell **104**: 829–838.

Iqbal, K., Jin, S.-G., Pfeifer, G.P., and Szabó, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci USA **108**: 3642–3647.

Johnson, D.R. (1974). Hairpin-tail: a case of post-reductional gene action in the mouse egg. Genetics **76**: 795–805.

Jullien, P.E. and Berger, F. (2009). Gamete-specific epigenetic mechanisms shape genomic imprinting. Curr Opin Plant Biol **12**: 637–642.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase *Dnmt3a* in paternal and maternal imprinting. Nature **429**: 900–903.

Kawahara, M., Wu, Q., Takahashi, N., Morita, S., Yamada, K., Ito, M., Ferguson-Smith, A.C., and Kono, T. (2007). High-frequency generation of viable mice from engineered bi-maternal embryos. Nat Biotech **25**: 1045–1050.

Kermicle, J.L. (1970). Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. Genetics **66**: 69–85.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. Science **303**: 521–523.

Kono, T., Obata, Y., Wu, Q., Niwa, K., Ono, Y., Yamamoto, Y., Park, E.S., Seo, J.-S., and Ogawa, H. (2004). Birth of parthenogenetic mice that can develop to adulthood. Nature **428**: 860–864.

Lau, M.M., Stewart, C.E., Liu, Z., Bhatt, H., Rotwein, P., and Stewart, C.L. (1994). Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. Genes Dev **8**: 2953–2963.

Lee, J.T. (2011). Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. Nat Rev Mol Cell Biol **12**: 815–826.

Lefebvre, L., Viville, S., Barton, S.C., Ishino, F., Keverne, E.B., and Surani, M.A. (1998). Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene *Mest*. Nat Genet **20**: 163–169.

Lercher, M.J. and Hurst, L.D. (2003). Imprinted chromosomal regions of the human genome have unusually high recombination rates. Genetics **165**: 1629–1632.

Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., Dean, W., Walter, J., Higgins, M., Feil, R., and Reik, W. (2004). Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nat Genet **36**: 1291–1295.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature **366**: 362–365.

Li, L., Keverne, E.B., Aparicio, S.A., Ishino, F., Barton, S.C., and Surani, M.A. (1999). Regulation of maternal behavior and offspring growth by paternally expressed *Peg3*. Science **284**: 330–333.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. Dev Cell **15**: 547–557.

Lin, S.-P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J., and Ferguson-Smith, A.C. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1-Gtl2* imprinted cluster on mouse chromosome 12. Nat Genet **35**: 97–102.

Ludwig, T., Eggenschwiler, J., Fisher, P., D'Ercole, A.J., Davenport, M.L., and Efstratiadis, A. (1996). Mouse mutants lacking the type 2 IGF receptor (IGF2R) are rescued from perinatal lethality in *Igf2* and *Igf1r* null backgrounds. Dev Biol **177**: 517–535.

Lyle, R., Watanabe, D., te Vruchte, D., Lerchner, W., Smrzka, O.W., Wutz, A., Schageman, J., Hahner, L., Davies, C., and Barlow, D.P. (2000). The imprinted antisense RNA at the *Igf2r* locus overlaps but does not imprint *Mas1*. Nat Genet **25**: 19–21.

Mancini-Dinardo, D., Steele, S.J.S., Ingram, R.S., and Tilghman, S.M. (2003). A differentially methylated region within the gene *Kcnq1* functions as an imprinted promoter and silencer. Hum Mol Genet **12**: 283–294.

Mancini-Dinardo, D., Steele, S.J.S., Levorse, J.M., Ingram, R.S., and Tilghman, S.M. (2006). Elongation of the *Kcnq1ot1* transcript is required for genomic imprinting of neighboring genes. Genes Dev **20**: 1268–1282.

McGrath, J. and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell **37**: 179–183.

Moore, T. and Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet **7**: 45–49.

Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. (2005). Epigenetic reprogramming in mammals. Hum Mol Genet **14 Spec No**: R47–58.

Murrell, A., Heeson, S., and Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. Nat Genet **36**: 889–893.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science **322**: 1717–1720.

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., Tanaka, S., Shiota, K., and Nakano, T. (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol **9**: 64–71.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., Cheng, X., and Bestor, T.H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature **448**: 714–717.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell **32**: 232–246.

Pàldi, A., Gyapay, G., and Jami, J. (1995). Imprinted chromosomal regions of the human genome display sex-specific meiotic recombination frequencies. Curr Biol **5**: 1030–1035.

Raissig, M.T., Baroux, C., and Grossniklaus, U. (2011). Regulation and Flexibility of Genomic Imprinting during Seed Development. Plant Cell **23**: 16–26.

Redrup, L., Branco, M.R., Perdeaux, E.R., Krueger, C., Lewis, A., Santos, F., Nagano, T., Cobb, B.S., Fraser, P., and Reik, W. (2009). The long noncoding RNA *Kcnq1ot1* organises a lineage-specific nuclear domain for epigenetic gene silencing. Development **136**: 525–530.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature **447**: 425–432.

Reik, W., Constância, M., Fowden, A., Anderson, N., Dean, W., Ferguson-Smith, A., Tycko, B., and Sibley, C. (2003). Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. J Physiol **547**: 35–44.

Reik, W. and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat Rev Genet **2**: 21–32.

Renfree, M.B., Hore, T.A., Shaw, G., Graves, J.A.M., and Pask, A.J. (2009). Evolution of genomic imprinting: insights from marsupials and monotremes. Annu Rev Genomics Hum Genet **10**: 241–262.

Robinson, W.P. and Lalande, M. (1995). Sex-specific meiotic recombination in the Prader--Willi/Angelman syndrome imprinted region. Hum Mol Genet **4**: 801–806.

Sandovici, I., Kassovska-Bratinova, S., Vaughan, J.E., Stewart, R., Leppert, M., and Sapienza, C. (2006). Human imprinted chromosomal regions are historical hot-spots of recombination. PLoS Genet **2**: e101.

Sharman, G.B. (1971). Late DNA replication in the paternally derived X chromosome of female kangaroos. Nature **230**: 231–232.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding *Air* RNA is required for silencing autosomal imprinted genes. Nature **415**: 810–813.

Smilinich, N.J., Day, C.D., Fitzpatrick, G.V., Caldwell, G.M., Lossie, A.C., Cooper, P.R., Smallwood, A.C., Joyce, J.A., Schofield, P.N., Reik, W., Nicholls, R.D., Weksberg, R., Driscoll, D.J., Maher, E.R., Shows, T.B., and Higgins, M.J. (1999). A maternally methylated CpG island in *KvLQT1* is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. Proc Natl Acad Sci USA **96**: 8064–8069.

Smits, G., Mungall, A.J., Griffiths-Jones, S., Smith, P., Beury, D., Matthews, L., Rogers, J., Pask, A.J., Shaw, G., VandeBerg, J.L., McCarrey, J.R., consortium, S., Renfree, M.B., Reik, W., and Dunham, I. (2008). Conservation of the *H19* noncoding RNA and *H19-IGF2* imprinting mechanism in therians. Nat Genet **40**: 971–976.

Spillane, C., Baroux, C., Escobar-Restrepo, J.M., Page, D.R., Laoueille, S., and Grossniklaus, U. (2004). Transposons and tandem repeats are not involved in the control of genomic imprinting at the *MEDEA* locus in *Arabidopsis*. Cold Spring Harb Symp Quant Biol **69**: 465–475.

Surani, M.A., Barton, S.C., and Norris, M.L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature **308**: 548–550.

Sutcliffe, J.S., Nakao, M., Christian, S., Orstavik, K.H., Tommerup, N., Ledbetter, D.H., and Beaudet, A.L. (1994). Deletions of a differentially methylated CpG island at the *SNRPN* gene define a putative imprinting control region. Nat Genet **8**: 52–58.

Suzuki, S., Ono, R., Narita, T., Pask, A.J., Shaw, G., Wang, C., Kohda, T., Alsop, A.E., Marshall Graves, J.A., Kohara, Y., Ishino, F., Renfree, M.B., and Kaneko-Ishino, T. (2007). Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. PLoS Genet **3**: e55.

Szabó, P., Tang, S.H., Rentsendorj, A., Pfeifer, G.P., and Mann, J.R. (2000). Maternal-specific footprints at putative *CTCF* sites in the *H19* imprinting control region give evidence for insulator function. Curr Biol **10**: 607–610.

Takagi, N. and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature **256**: 640–642.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H.F.M. (2008). *Polycomb* group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev Cell **15**: 668–679.

Thorvaldsen, J.L., Duran, K.L., and Bartolomei, M.S. (1998). Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. Genes Dev **12**: 3693–3702.

Tunster, S.J., Tycko, B., and John, R.M. (2010). The imprinted *Phlda2* gene regulates extraembryonic energy stores. Mol Cell Biol **30**: 295–306.

Tycko, B. and Morison, I.M. (2002). Physiological functions of imprinted genes. J Cell Physiol **192**: 245–258.

Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y., and Feil, R. (2004). Imprinting along the *Kcnq1* domain on mouse chromosome 7 involves repressive histone methylation and recruitment of *Polycomb* group complexes. Nat Genet **36**: 1296–1300.

Varmuza, S. and Mann, M. (1994). Genomic imprinting--defusing the ovarian time bomb. Trends Genet **10**: 118–123.

de Villena, F.P.-M., de la Casa-Esperón, E., and Sapienza, C. (2000). Natural selection and the function of genome imprinting: beyond the silenced minority. Trends Genet **16**: 573–579.

Williamson, C.M., Turner, M.D., Ball, S.T., Nottingham, W.T., Glenister, P., Fray, M., Tymowska-Lalanne, Z., Plagge, A., Powles-Glover, N., Kelsey, G., Maconochie, M., and Peters, J. (2006). Identification of an imprinting control region affecting the expression of all transcripts in the *Gnas* cluster. Nat Genet **38**: 350–355.

Wolf, J.B. and Hager, R. (2006). A maternal-offspring coadaptation theory for the evolution of genomic imprinting. PLoS Biol **4**: e380.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C. (2011). High-Resolution Analysis of Parent-of-Origin Allelic Expression in the *Arabidopsis* Endosperm. PLoS Genet **7**: e1002126.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun **2**: 241.

Wutz, A., Smrzka, O.W., Schweifer, N., Schellander, K., Wagner, E.F., and Barlow, D.P. (1997). Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. Nature **389**: 745–749.

Yan, Y., Frisén, J., Lee, M.H., Massagué, J., and Barbacid, M. (1997). Ablation of the CDK inhibitor *p57Kip2* results in increased apoptosis and delayed differentiation during mouse development. Genes Dev **11**: 973–983.

INTRODUCTION
CHAPTER 2

GENOMIC IMPRINTING IN FLOWERING PLANTS

NOTE

The following review is published as **Raissig, M.T., Baroux, C., and Grossniklaus, U**. (2011). Regulation and Flexibility of Genomic Imprinting during Seed Development. Plant Cell **23**: 16–26. It is under the Copyright of the American Society of Plant Biology and can be found under www.plantcell.org.

MTR wrote and revised the review and conceived and designed the figures. CB and UG critically read, corrected and modified the article. Supplemental information of the review can be found in Appendix A1.

The Plant Cell, Vol. 23: 16–26, January 2011, www.plantcell.org

REVIEW

Regulation and Flexibility of Genomic Imprinting during Seed Development [W]

Michael T. Raissig, Célia Baroux, and Ueli Grossniklaus[1]

Institute of Plant Biology and Zürich-Basel Plant Science Center, University of Zürich, CH-8008 Zurich, Switzerland

Genomic imprinting results in monoallelic gene expression in a parent-of-origin–dependent manner. It is achieved by the differential epigenetic marking of parental alleles. Over the past decade, studies in the model systems *Arabidopsis thaliana* and maize (*Zea mays*) have shown a strong correlation between silent or active states with epigenetic marks, such as DNA methylation and histone modifications, but the nature of the primary imprint has not been clearly established for all imprinted genes. Phenotypes and expression patterns of imprinted genes have fueled the perception that genomic imprinting is specific to the endosperm, a seed tissue that does not contribute to the next generation. However, several lines of evidence suggest a potential role for imprinting in the embryo, raising questions as to how imprints are erased and reset from one generation to the next. Imprinting regulation in flowering plants shows striking similarities, but also some important differences, compared with the mechanisms of imprinting described in mammals. For example, some imprinted genes are involved in seed growth and viability in plants, which is similar in mammals, where imprinted gene regulation is essential for embryonic development. However, it seems to be more flexible in plants, as imprinting requirements can be bypassed to allow the development of clonal offspring in apomicts.

INTRODUCTION

The diploid phase of the life cycle is dominant in the majority of multicellular organisms. As a result, deleterious recessive mutations are masked during the diploid phase (Otto and Goldstein, 1992), implying an equally important role of both parental alleles. While this statement is true for most genes in the genome, imprinted genes represent an exception because only one parental allele is expressed while the other remains silent. Several theories have been proposed to explain the evolution of this epigenetic phenomenon, which is found in organisms as evolutionarily divergent as mammals and flowering plants (reviewed in Haig and Westoby, 1989; Hurst and McVean, 1998; Baroux et al., 2002; Gutierrez-Marcos et al., 2003; Wilkins and Haig, 2003; Feil and Berger, 2007; Moore and Mills, 2008). Because genomic imprinting often results in parent-of-origin–specific effects on the growth of the embryo and extra-embryonic tissues in mammals and plants, the parental conflict theory provides one of the most widely accepted explanations.

Both mammals and flowering plants use a common reproductive strategy (i.e., they share a placental habit). The embryo is embedded and nourished by sexually derived, extra-embryonic tissues: the placenta in mammals and the endosperm in plants. The mammalian placenta and embryo are derived from the same fertilization event and embryonic cells partition early into an inner cell mass, which forms the embryo, and the trophectoderm, which will participate in the formation of the placenta. By contrast, the plant embryo and endosperm derive from two distinct fertilization events involving two female gametes, the egg and the central cell, and two sperm cells. The central cell is homodiploid and thus contributes two maternal genomes to the triploid endosperm, whereas only one genome is of paternal origin. This genetic peculiarity is of importance for seed formation and considerably complicates the interpretation of parent-of-origin–specific effects during seed development (reviewed in Birchler, 1993; Spillane et al., 2002; Dilkes and Comai, 2004; von Wangenheim and Peterson, 2004).

Genomic imprinting is conveyed by an epigenetic, parent-of-origin–specific mark (the imprint), which leads to the differential expression of the parental alleles. In mammals, where genomic imprinting is best understood, the imprint is set during gametogenesis, interpreted and maintained during development, and erased and reset in the germ line for the next generation. Genomic imprinting seems to irreversibly set the epigenetic state of certain parental alleles during gametogenesis in animals. As a result, it prevents normal development of gyno- and androgenotes, which carry two maternal or paternal genomes, respectively. Genomic imprinting in plants shares some common principles of regulation with animals and shows some versatility. For instance, it is modulated by parental genomic dosage (Erilova et al., 2009; Tiwari et al., 2010) and might be abrogated in successful hybridization events (Josefsson et al., 2006; Walia et al., 2009), reminiscent of the disruption of genomic imprinting at some loci in interspecific crosses of rodents in the genus *Peromyscus* (Vrana et al., 1998). Lastly, apomixis, the asexual reproduction through seeds without paternal contribution, likely requires a bypass of genomic imprinting at least in some species (Koltunow and Grossniklaus, 2003; Grossniklaus, 2009).

[1] Address correspondence to grossnik@botinst.uzh.ch.
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.110.081018

GENOMIC IMPRINTING AND THE INTRAGENOMIC PARENTAL CONFLICT THEORY

Reciprocal crosses of maize (*Zea mays*) varieties with differently colored kernels led to the discovery of gene-specific imprinting. In 1970, Kermicle demonstrated that full kernel pigmentation depends on maternal inheritance of the *R1* gene, which regulates anthocyanin biosynthesis in the endosperm. By contrast, a mottled pigmentation results when *R1* is inherited paternally (Kermicle, 1970). Through a series of elegant genetic experiments, Kermicle could show that this difference in phenotype is due neither to cytoplasmic inheritance nor to a dosage effect in the endosperm but depends solely on the parental origin of *R1*. The implication of this work was not widely recognized at the time. However, over a decade later, the importance of genomic imprinting was highlighted by nuclear transfer assays in mouse oocytes, showing that both parental genomes are required for normal development of the embryo and, thus, for successful reproduction. Embryonic development cannot be completed if an enucleated egg cell receives either two female (gynogenote, 2m:0p) or two male pronuclei (androgenote, 0m:2p) (Barton et al., 1984; McGrath and Solter, 1984; Surani et al., 1984). Because uniparental disomies of some chromosomes develop normally while some maternal duplications cannot rescue the corresponding paternal deficiencies and vice versa, Cattanach and Kirk (1985) concluded that the entire genome is not subject to parental effects but that imprinted loci reside in specific chromosomal regions. This suggests that the two parental genomes are not equivalent and one genome of each parent is required to complete development, with abnormal expression of imprinted genes underlying the incomplete development of andro- and gynogenotes. Effects of stored components in the cytoplasm of the egg or dosage effects can be excluded, since cytoplasm and parental dosage, except for that of imprinted genes, is the same in both cases.

In placental mammals, mutations in many imprinted genes cause placenta and embryo growth defects in a parent-of-origin–specific manner (DeChiara et al., 1991; Lau et al., 1994; Tycko and Morison, 2002). Similarly in flowering plants, mutations in the imprinted *Arabidopsis Polycomb* group genes *MEDEA* (*MEA*) and *FERTILISATION-INDEPENDENT SEED2* (*FIS2*) induce proliferation defects during seed development (Grossniklaus et al., 1998; Kinoshita et al., 1999; Kiyosue et al., 1999; Vielle-Calzada et al., 1999; Luo et al., 2000; Ingouff et al., 2005b). These defects are consistent with a role of genomic imprinting in a parental conflict over resource allocation from mother to offspring. In mammals and seed plants, postfertilization nutrient provision is at the cost of the mother only, a situation that reduces the resources available for future offspring. Thus, in polygamous organisms with a placental habit, maternally expressed genes are expected to favor parsimonious distribution of nutrients and to antagonize paternally expressed growth factors. In this scenario, growth-restricting genes evolved to be under maternal control, whereas growth-promoting genes are expected to be under paternal control (Haig and Westoby, 1989; Haig and Graham, 1991; Moore and Haig, 1991).

In maize, deletions of specific chromosomal arms reduce kernel size and viability when inherited paternally. These defects cannot be rescued with a higher, maternal dosage of the missing chromosomal arm (reviewed in Kermicle and Alleman, 1990). Thus, genes located on the deleted chromosomal arm may only be expressed from one parent and cannot be rescued by inheriting a compensating dose from the other parent. In addition, increasing the dosage of one parent in *Arabidopsis thaliana* affects endosperm growth and, consequently, seed size. Normal endosperm development requires a maternal:paternal genome ratio of 2:1 (2m:1p) in many species (Lin, 1984). In some *Arabidopsis* accessions, however, deviations of this ratio are tolerated, and crosses between different ploidies can produce viable seeds. Consistent with the predictions made from the parental conflict theory, a cross between a tetraploid mother and diploid father (4n $\times$ 2n) produces small seeds, whereas the reciprocal cross (2n $\times$ 4n) produces larger seeds (Scott et al., 1998). However, these parent-of-origin effects in seeds derived from interploidy crosses are more complex to interpret than the nuclear transfer experiments in mice. Although imprinting of growth regulators likely contribute to the phenotypes, these experiments do not allow precise differentiation from other effects, such as the dosage of cytoplasmically inherited products (Birchler, 1993; Dilkes and Comai, 2004; von Wangenheim and Peterson, 2004) and the role of surrounding maternal tissues (Garcia et al., 2005; Dilkes et al., 2008). The effects on seed size thus likely result from an interplay between cytoplasmic effects, dosage effects, and genomic imprinting.

Alternative theories have been proposed to explain the evolution of genomic imprinting and are reviewed elsewhere (Garnier et al., 2008; Moore and Mills, 2008). One of these proposes that genomic imprinting evolved to prevent parthenogenesis, explaining why gynogenotes and androgenotes fail to complete development (Barton et al., 1984; McGrath and Solter, 1984; Surani et al., 1984; Solter, 1988; Varmuza and Mann, 1994). Consistent with this theory, screens for autonomous endosperm development in *Arabidopsis* revealed, among others, two imprinted genes, namely, *MEA* and *FIS2* (Chaudhury et al., 1997; Kiyosue et al., 1999). However, this phenotype is also observed in mutants affecting nonimprinted *Polycomb* group genes, and mutation or downregulation of these does not seem to be responsible for parthenogenesis in apomictic *Hieracium* (Rodrigues et al., 2008).

It was also hypothesized that genomic imprinting evolved as a consequence of repressing foreign DNA, namely, transposable elements (TEs) (Barlow, 1993). In *Arabidopsis*, methylated promoter regions of *MEA* and *FWA* coincide with TEs (Lippman et al., 2004; Spillane et al., 2004; Gehring et al., 2006; Kinoshita et al., 2007), but the presence of TEs is not necessarily causally related to imprinted expression. While the SINE-related sequence in the *FWA* control region seems to control imprinted *FWA* expression (Kinoshita et al., 2007), the helitron in the *MEA* promoter is dispensable for imprinting (Spillane et al., 2004), as are the direct repeats downstream of *MEA*, despite being differentially methylated in the endosperm (Spillane et al., 2004; Gehring et al., 2006). The finding that TEs are hypomethylated in the genome of *Arabidopsis* endosperm compared with that of the embryo (Gehring et al., 2009; Hsieh et al., 2009) indicates a potential involvement of TE-related sequences in regulating imprinted loci.

DEFINING IMPRINTED LOCI: MEETING THE STANDARDS

Imprinting is defined by the differential expression of parental alleles in the same nucleus. Typically, mammalian imprinted genes have been identified using allele-specific measurements in somatic cells of either transcript levels, DNA methylation, reporter gene activity, or mutant analyses following reciprocal crosses. Imprinted and potentially imprinted genes in flowering plants have been identified using similar approaches (Table 1). Most of the imprinted and potentially imprinted genes have been qualified as imprinted based on differential levels of parental transcripts and/or detection of a reporter gene when inherited through one parent but not the other. Imprinting (de novo monoallelic expression after fertilization) but also cytoplasmic inheritance of maternal or paternal transcripts can explain these observations (Grossniklaus et al., 1998; Vielle-Calzada et al., 2000; Bayer et al., 2009). Thus, an analysis of steady state levels of parental transcripts is only sufficient to demonstrate imprinting if the candidate gene is not expressed prior to fertilization in the gametes, as is the case for the *Arabidopsis* genes *PHERES1* (*PHE1*) and *FORMIN HOMOLOG5* (*FH5*) and, in the embryo, for the maize gene *Maternally expressed in embryo1* (*Mee1*) (Köhler et al., 2005; Fitz Gerald et al., 2009; Jahnke and Scholten, 2009; Wuest et al., 2010). Alternatively, parent-of-origin–specific expression that increases after fertilization provides evidence for genomic imprinting in addition to parental transcripts potentially inherited from the gametes, whose abundance cannot increase. However, such analyses require quantitative methods and, importantly, proper internal standards that are stably expressed during development in the fertilization products only. This is because carpel cell number is stable during silique development and growth solely depends on cell expansion (Vivian-Smith and Koltunow, 1999). In addition, integument cells proliferate only early in seed development, whereas later on, cell elongation accounts for integument growth (Garcia et al., 2005). Thus, internal standards expressed in all cells of the developing silique do not allow a precise quantitation of mRNA levels. Reliable quantitation requires either standards that are expressed in the fertilization products only or experiments using isolated tissues, the latter being technically demanding at early stages. Thus, monoallelic de novo transcription of candidate imprinted loci is best demonstrated by (1) nascent transcript detection using RNA in situ hybridization, as shown for the imprinted *MEA* gene in the endosperm (Vielle-Calzada et al., 1999), (2) absence of expression prior to fertilization (Köhler et al., 2005; Fitz Gerald et al., 2009; Jahnke and Scholten, 2009; Wuest et al., 2010), or (3) nuclear run-off transcription assays, which are not possible at early stages due to the limited material available. To remain consistent with previous literature, we discuss both imprinted and potentially imprinted genes (i.e., for which monoallelic expression has not been demonstrated unambiguously) in the following sections.

Furthermore, defining the primary imprint remains a challenging task that requires epigenetic profiling of the candidate loci in isolated male and female gametes, as first performed in maize for *Fie1* and *Fie2* (Gutiérrez-Marcos et al., 2006). To date, a functional verification showing that the identified mark is indeed necessary for imprinted expression has only been performed for the *FWA* locus in *Arabidopsis* (Kinoshita et al., 2007). Importantly, defining the epigenetic profile several days after pollination (DAP) is not sufficient because epigenetic marks can be highly dynamic (Gutiérrez-Marcos et al., 2006; Jahnke and Scholten, 2009) and the marks maintaining imprinting may be different from the primary imprint(s). Although challenging, recent progress in isolating plant gametes and zygotes (Dresselhaus et al., 1999; Engel et al., 2003; Gutiérrez-Marcos et al., 2006; Ning et al., 2006; Hermon et al., 2007; Wuest et al., 2010) should allow a better characterization of the differential epigenetic states of imprinted alleles.

MOLECULAR AND CELLULAR FUNCTIONS OF IMPRINTED GENES DURING SEED DEVELOPMENT

To date, 11 genes in *Arabidopsis* and 11 genes in maize have been reported as imprinted or potentially imprinted (Table 1). They have been identified on the basis of parent-of-origin–specific effects on seed development (Kermicle, 1970; Grossniklaus et al., 1998; Kinoshita et al., 1999; Vielle-Calzada et al., 1999; Luo et al., 2000), differential transcript levels in interploidy or interaccession crosses (Chaudhuri and Messing, 1994; Lund et al., 1995a, 1995b; Gutiérrez-Marcos et al., 2004; Tiwari et al., 2008; Jahnke and Scholten, 2009), or differential DNA methylation levels between embryo and endosperm (Gehring et al., 2009). The latter is not an imprinting criterium per se but may strongly favor the identification of imprinted genes expressed in the endosperm. This study led to the identification of 50 candidate imprinted loci (Gehring et al., 2009), but future investigations are needed to establish whether or not they are indeed regulated by genomic imprinting.

Imprinted genes encode a wide range of molecular functions, ranging from the regulation of pigmentation, protein storage, transcriptional regulation, chromatin modification, and cytoskeletal function to mRNA regulation (Table 1). For instance, five recently described potentially imprinted genes, for which only transcripts from one parental allele were detected in the endosperm (Gehring et al., 2009), encode transcription factors of the homeodomain and MYB class (Table 1). The function of these genes is currently unknown, and future studies will show whether they have parent-of-origin–specific roles during endosperm development. Because most imprinted loci were identified recently, little is known about their role during development, except for four genes in *Arabidopsis*. Mutations in either of the two *Polycomb* group genes *MEA* and *FIS2* confer maternal effects on seed development, displaying proliferation defects in the endosperm with and without fertilization (Chaudhury et al., 1997; Grossniklaus and Vielle-Calzada, 1998; Grossniklaus et al., 1998; Kiyosue et al., 1999; Ingouff et al., 2005b). The seeds containing embryo and endosperm derived from mutant gametes eventually abort. Their development is delayed such that embryos derived from mutant eggs only reach the late heart or sometimes torpedo stage, while their wild-type siblings complete embryogenesis (Chaudhury et al., 1997; Grossniklaus et al., 1998; Kiyosue et al., 1999; Ingouff et al., 2005b). When comparing wild-type and mutant embryos at the same developmental stage, however, the mutant embryos show overproliferation, leading to the formation of several extra cell layers (Grossniklaus et al., 1998). Similarly, mutants in the *Arabidopsis* genes *FH5* and

Table 1. Imprinted Genes in *Arabidopsis* and Maize

	Function			Parent-of-Origin–Specific Effect for				Reference
Gene	Encoded Protein	Cellular Function	De Novo Transcription	mRNA Levels	Reporter Activity	Epigenetic Mark	Mutant Phenotype	For Parent-of-Origin–Specific Expression
Arabidopsis								
MEA	*Pc*G complex protein	Cell proliferation (embryo, endosperm)	Yes	Yes	Yes	H3K27me3, DNA-me	Seed abortion	Kinoshita et al. (1999); Vielle-Calzada et al. (1999)
FWA	HD-ZIP transcription factor	Flowering time regulation[a]	n.d.	Yes	n.d.	DNA-me	No phenotype	Kinoshita et al. (2004); Jullien et al. (2006a)
PHE1	MADS box transcription factor	Not known	Yes[b]	Yes	Yes	H3K27me3, DNA-me	No phenotype	Köhler et al. (2003, 2005); Makarevich et al. (2008)
FIS2	*Pc*G complex protein	Cell Proliferation (embryo, endosperm)	n.d.	Yes	Yes	DNA-me[c]	Seed abortion	Jullien et al. (2006a, 2008); Luo et al. (2000)
MPC	Poly(A) binding C-terminal domain	Not known	n.d.	Yes	Yes	DNA-me[c]	Small seeds, abnormal embryo	Tiwari et al. (2008)
HDG9	HD-ZIP transcription factor	Not known	n.d.	Yes	n.d.	DNA-me	Not known	Gehring et al. (2009)
HDG8	HD-ZIP transcription factor	Not known	n.d.	Yes	n.d.	DNA-me	Not known	Gehring et al. (2009)
HDG3	HD-ZIP transcription factor	Not known	n.d.	Yes	n.d.	DNA-me	Not known	Gehring et al. (2009)
MYB3R2	MYB transcription factor	Not known	n.d.	Yes	n.d.	DNA-me	Not known	Gehring et al. (2009)
AT5G62110	Homeodomain-like protein	Not known	n.d.	Yes	n.d.	DNA-me	Not known	Gehring et al. (2009)
FH5	Formin homolog	Morphogenesis, cellularization (endosperm)	Yes[d]	Yes	Yes	H3K27me3[c]	Endosperm defects[c]	Ingouff et al. (2005a); Fitz Gerald et al. (2009)
Maize (locus-specific imprinted genes)								
Fie1	Homolog of *AthFIE*	Not known	Yes[b]	Yes	n.d.	DNA-me, H3K27me3, H3/H4-Ac	Not known	Danilevskaya et al. (2003); Gutiérrez-Marcos et al. (2006); Hermon et al. (2007); Haun and Springer (2008)
Fie2	Homolog of *AthFIE*	Not known	n.d.	Yes	n.d.	DNA-me	Not known	Danilevskaya et al. (2003); Gutiérrez-Marcos et al. (2006); Hermon et al. (2007)
Nrp1	Putative transcription factor	Not known	Yes[b]	Yes	n.d.	DNA-me, H3K27me3, H3/H4-Ac	Not known	Guo et al. (2003); Haun and Springer (2008)
Peg1	Not known	Not known	n.d.	Yes	n.d.	Not known	Not known	Gutierrez-Marcos et al. (2003)
Meg1	Small Cys-rich polypeptide	Nutrient transfer?	Yes[b]	Yes	Yes	DNA-me	Not known	Gutiérrez-Marcos et al. (2004)
Mez1	Homolog of *MEA*	Not known	n.d.	Yes	n.d.	DNA-me, H3K27me3, H3/H4-Ac	Not known	Haun et al. (2007); Haun and Springer (2008)
Mee1	Unknown protein	Not known	Yes[e]	Yes	n.d.	DNA-me	Not known	Jahnke and Scholten (2009)
Maize (allele-specific imprinted genes)								
α-Tubulin	Tubulin homolog	Cytoskeleton	n.d.	Yes	n.d.	DNA-me	Not known	Lund et al. (1995b)
Zein	Zein protein	Storage	n.d.	Yes	n.d.	DNA-me	Not known	Lund et al. (1995a)
R gene	Myc-like transcription factor	Anthocyanin pigmentation	n.d.	n.d.	n.d.	Not known	Not known	Kermicle (1970); Ludwig et al. (1989)
Dzr-1	Not known	Zein regulation (posttranscriptional)	n.d.	Yes	n.d.	Not known	Not known	Chaudhuri & Messing (1994)

The table displays the imprinted and potentially imprinted genes discovered in plants to date. Molecular and cellular functions and investigated parent-of-origin–specific effects, such as de novo gene expression after fertilization, steady state mRNA levels, and imprinted reporter gene expression are listed if known. In addition, phenotypes of the associated mutants and the associated epigenetic marks are specified. Genes in regular font are paternally expressed, whereas genes in boldface are maternally active. n.d., not determined.

[a]Overexpression phenotype; function in seed development unknown.

[b]Expression after fertilization only.

[c]Epigenetic mark/phenotype shown but not specifically in a parent-of-origin–specific manner.

[d]Expression after fertilization only as no expression was detected in the central cell (Wuest et al., 2010).

[e]De novo expression in the embryo.

MATERNALLY EXPRESSED PAB C-TERMINAL (*MPC*) display defects in endosperm development. However, they have distinct molecular functions in cytoskeleton and mRNA biology, respectively (Ingouff et al., 2005a; Tiwari et al., 2008; Fitz Gerald et al., 2009). Furthermore, *Maternally expressed gene1* (*Meg1*), an imprinted gene in maize, is specifically expressed in the transfer cells (Gutiérrez-Marcos et al., 2004; Table 1). These cells are involved in nutrient transfer from the maternal tissues to the seed, suggesting a nutrition-related function of *Meg1*, a proposal that awaits demonstration. Altogether, the known or predicted functions of imprinted or potentially imprinted genes indicate a role in endosperm growth and nutrient transfer to the seed, consistent with the proposed role of genomic imprinting in mediating parent-of-origin–specific effects on resource allocation.

In contrast with the endosperm, very little is known about the role of imprinted genes during embryo development, although embryonic phenotypes were originally described for mutants of the *fis* class and in *MPC* RNA interference lines (Ohad et al., 1996; Chaudhury et al., 1997; Grossniklaus et al., 1998; Tiwari et al., 2008). For instance, embryos lacking maternal *MEA* function overproliferate (Grossniklaus et al., 1998; Luo et al., 2000), a phenotype similar to the embryo overgrowth resulting from, for example, a mutation in the mouse imprinted gene *Insulin-like growth factor type 2 receptor* (Barlow et al., 1991; Lau et al., 1994). *MEA* is expressed in the egg cell and/or early embryo, as shown by RNA in situ hybridization (Vielle-Calzada et al., 1999; Spillane et al., 2007), reporter gene analyses (Luo et al., 2000; Spillane et al., 2004, 2007; Figure 1A), and RT-PCR on isolated embryos (Figure 1B). Whether the embryonic expression of *MEA* is imprinted remains a matter of debate. On the one hand, qualitative RT-PCR using a natural polymorphism between the Landsberg *erecta* and the RLD accession for allele-specific detection of the parental *MEA* transcripts argues for biallelic expression at 4 to 8 DAP (Kinoshita et al., 1999; Gehring et al., 2006; Erilova et al., 2009). On the other hand, paternal *MEA* transcripts were not detected in RT-PCR experiments at 2.3 DAP (Vielle-Calzada et al., 1999) nor in a time-course experiment of developing seeds using a quantitative real-time RT-PCR assay based on a polymorphism between the *mea-2* and the wild-type allele of the same accession (Baroux et al., 2006; Figure 1C). Possibly, the discrepancies between these studies reflect the different accessions used because it is known that hybridization events can interfere with genomic imprinting (Josefsson et al., 2006; Walia et al., 2009).

A role for genomic imprinting, or at least for imprinted genes, in embryogenesis has often been dismissed (Gehring et al., 2004; Jahnke and Scholten, 2009; Jullien and Berger, 2009). The embryonic phenotype and expression pattern of *MEA* indicates yet unknown roles for *MEA*, possibly in controlling growth regulators in the embryo. Similarly, functional studies will shed more light on the embryonic role of *PHE1*, which is expressed and regulated by the *Polycomb* Repressive Complex 2 containing MEA and other FIS class proteins (FIS-PRC2) in both embryo and endosperm (Köhler et al., 2003, 2005) and the *Mee1* gene in maize, which also shows imprinted expression in the embryo (Jahnke and Scholten, 2009).

In conclusion, the best characterized imprinted or potentially imprinted plant genes share a role in endosperm development or are at least preferentially expressed in this tissue. When mutated, some of these genes show endosperm growth abnormalities consistent with a proposed role of genomic imprinting in parental conflicts over resource allocation. Nevertheless, additional studies are required to address the function of imprinted genes in the endosperm and, particularly, in the embryo.

REGULATION OF GENOMIC IMPRINTING: WHAT DO WE KNOW; WHERE DO WE GO?

Regulation of genomic imprinting in mammals is complex, and several epigenetic mechanisms involving DNA methylation, histone modifications, and noncoding RNAs are recruited to define

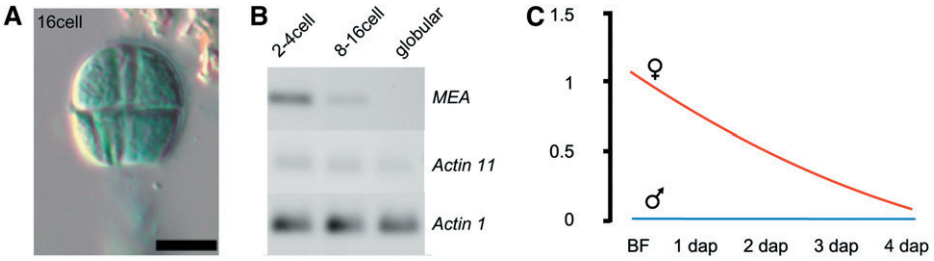


Figure 1. Expression of *MEA* in the *Arabidopsis* Embryo.

(A) Expression of a MEA:GUS reporter gene expressing β-glucuronidase (GUS) under the control of a 3.8-kb promoter (Spillane et al., 2004) in an isolated 16-cell stage embryo. GUS stainings were performed as described by Baroux et al. (2006), except for using only 0.1 mM potassium hexacyanoferrate (II) and 0.1 mM potassium hexacyanoferrate (III) and incubating for 4 d at 37°C. As the half-life of GUS in early embryos is <10 h (R. Baskar and U. Grossniklaus, unpublished data), GUS expression of this stage indicates de novo expression.

(B) RT-PCR on isolated embryos amplifying *MEA* (34 cycles), as well as *ACTIN1* and *ACTIN11* (28 cycles each) as controls. Embryos were isolated in a buffer containing 1.6 units/μL RNase Out (Invitrogen) and 1 mM DTT. An inverted microscope was used to find the right stage and a glass capillary to collect the embryos. Conditions for RT-PCR and the primers detecting *MEA*, *ACTIN1*, and *ACTIN11* are described by Baroux et al. (2006). Images of isolated embryos and corresponding RNA profiles are shown in Supplemental Figure 1 online.

(C) Allele-specific quantitative PCR for *MEA* on siliques at different time points: before fertilization (BF) and 1 to 4 DAP. No paternal transcript can be detected, suggesting imprinted expression in both embryo and endosperm (Baroux et al., 2006).

the silent versus active state of parental alleles (reviewed in Ideraabdullah et al., 2008; Koerner and Barlow, 2010). Imprinting control elements (ICEs) are *cis*-regulatory sequences necessary and sufficient to confer imprinted expression. ICEs can function as promoters, enhancers, locus control regions, or insulator elements that control clusters of imprinted genes in a parent-of-origin–specific manner. ICEs are themselves subjected to epigenetic modifications and are usually differentially methylated (Barlow and Bartolomei, 2007). Our understanding of imprinting regulation in plants is much less profound than in animals. For instance, little is known about potential plant ICEs. Nevertheless, imprinting regulation in plants shows some parallels to the regulation of genomic imprinting in mammals. Over the last few years, variations of a predominant model have been developed to describe imprinting regulation in the *Arabidopsis* endosperm, which largely relies on the specific demethylation of maternal alleles in the central cell. However, we argue that additional models must be developed to take into account data from maize and to describe imprinting regulation in the embryo.

DNA Methylation: Establishing or Interpreting the Imprint?

Parental alleles of imprinted genes must be marked by a primary imprint inherited from the gametes and interpreted in the fertilization products. In principle, the imprint can be on the active or the silent allele, or both alleles can carry distinct marks. Identifying the imprint remains challenging, and nothing is known about the epigenetic status of imprinted loci in the gametes, except for a few loci in maize (Gutiérrez-Marcos et al., 2006; Jahnke and Scholten, 2009). Instead, genetic approaches in *Arabidopsis* have elucidated the control of imprinted expression, suggesting a fundamental role for DNA demethylation and FIS-PRC2 in imprinting regulation.

The maintenance DNA-methyltransferase *METHYLTRANSFERASE1* (*MET1*) and the DNA-glycosylase *DEMETER* (*DME*) act antagonistically to achieve monoallelic gene expression of *MEA*, *FWA*, *FIS2*, and *MPC* in *Arabidopsis* (Choi et al., 2002; Kinoshita et al., 2004; Gehring et al., 2006; Jullien et al., 2006a; Tiwari et al., 2008). The current model proposes that DNA methylation is actively removed by the action of DME (Choi et al., 2002; Gehring et al., 2006). This demethylation might be reinforced by passive loss in the central cell due to *MET1* down-regulation by RETINOBLASTOMA RELATED (RBR) and its interactor MULTICOPY SUPPRESSOR OF IRA1 (MSI1) (Johnston et al., 2008; Jullien et al., 2008). Consequently, gamete-specific demethylation activates the maternal alleles, leading to monoallelic expression of *MEA*, *FWA*, *FIS2*, and *MPC* after fertilization (Figure 2A). When *FWA*, *FIS2*, and *MPC* are inherited from a hypomethylated *met1* mutant father, their paternal alleles are expressed in the endosperm (Luo et al., 2000; Kinoshita et al., 2004; Tiwari et al., 2008), while a paternal *MEA* allele remains silent (Gehring et al., 2006; Jullien et al., 2006a). Thus, hypomethylation is sufficient to trigger biallelic expression of *FWA*, *FIS2*, and *MPC* but not of *MEA*. This suggests that default methylation by *MET1* and specific demethylation of the maternal alleles in the central cell by DME is required and sufficient for imprinting regulation at these loci except for *MEA* whose regulation is more complex. In fact, specific *MEA* reporter genes show imprinted expression independent of DME and MET1 activity (H. Wöhrmann and U. Grossniklaus, unpublished data), suggesting that additional, yet unknown, factors may be involved in imprinting regulation (Figure 2B).

Recent findings also implicate DNA methylation in the regulation of *PHE1*, a MADS box gene that is preferentially expressed from the paternal allele (Köhler et al., 2005) and plays a role in seed development (Köhler et al., 2003; Josefsson et al., 2006). A differentially methylated region has been identified outside the 3′ end of the *PHE1* coding region, and *PHE1* expression is reduced in *met1* mutant plants. Furthermore, a *PHE1* reporter gene is not expressed when inherited from a father deficient for DOMAINS REARRANGED METHYLASE1 (DRM1) and DRM2 (Makarevich et al., 2008), which encode de novo DNA methyltransferases (Cao and Jacobsen, 2002). Unlike other imprinted genes in plants, the hypermethylated allele is expressed, and DNA methylation seems to be involved in a mechanism leading to the activation of the paternal *PHE1* allele.

In maize, most of the candidate imprinted loci are differentially methylated in the endosperm, suggesting that DNA methylation plays a role in regulating imprinted gene expression in this species as well (Figure 2C, Table 1, and references therein). However, this asymmetry in DNA methylation may not be established prior to fertilization for all alleles. Indeed, *Fie2* is hypomethylated in both female gametes and the sperm, yet becomes transiently hypermethylated on the paternal alleles in the endosperm only. By contrast, *Fie1* is hypomethylated in the central cell but methylated in the egg and sperm cells. After fertilization, the paternal allele is specifically demethylated in the endosperm not correlating with its silent state (Gutiérrez-Marcos et al., 2006; Hermon et al., 2007). Similarly, DNA methylation at the *Mee1* locus is highly dynamic. *Mee1* shows imprinted expression in the embryo where the maternal allele is activated at 3 DAP, while it is already expressed prior to fertilization in the central cell and subsequently in the endosperm (Jahnke and Scholten, 2009). Maternal expression correlates with a hypomethylated state of *Mee1* in both fertilization products (Figure 2D). However, while the paternal methylation state was inherited from the sperm, the maternal state was not, with both female gametes showing hypermethylation. In the endosperm, the maternal alleles are demethylated after fertilization. In the embryo, they undergo demethylation in the zygote while initially remaining transcriptionally silent and then are remethylated during embryogenesis. These observations suggest that DNA methylation is not the primary imprint in these cases; rather, it may reinforce the transcriptionally active versus silent states. Differential DNA methylation presumably is established downstream of other gender-specific epigenetic marks, which remain to be discovered. Furthermore, these experiments clearly illustrate that DNA methylation is highly dynamic and not always correlated with expression. Due to its dynamics, it is not possible to infer the DNA methylation state in the gametes based on analyses performed at later stages of seed development.

Histone Modifications: Establishing and Maintaining Imprints?

The model that MET1 and DME antagonistically regulate genomic imprinting in plants was first developed for *MEA* (Choi

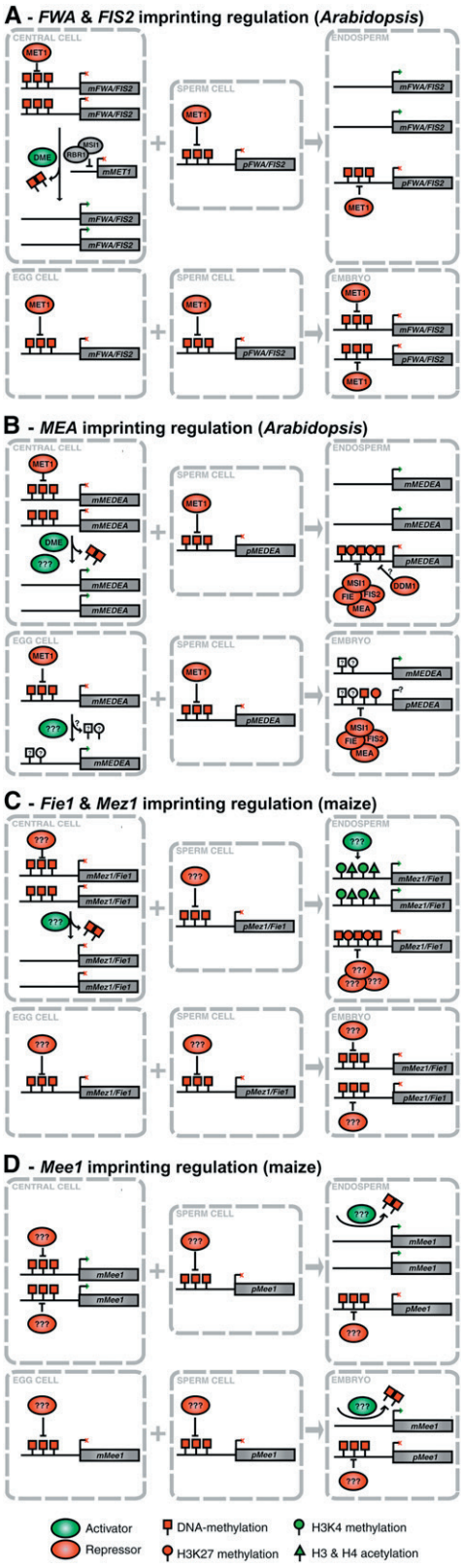


Figure 2. Imprinting Regulation of Maternally Expressed Genes in *Arabidopsis* and Maize.

et al., 2002) but turned out to be more complex, since the paternal *MEA* allele is not derepressed if inherited from a *met1* mutant father (Gehring et al., 2006; Jullien et al., 2006a). Thus, a different or additional epigenetic pathway must be involved to silence the paternal allele. FIS-PRC2 itself evidently maintains paternal alleles in a silent state since H3K27 methylation marks, which are dependent on PRC2 function, were found 5′ and 3′ of the *MEA* gene in siliques 7 DAP and leafs (Gehring et al., 2006; Jullien et al., 2006a). Importantly, seeds inheriting maternal *fie* or *mea* mutations show biparental *MEA* expression (Baroux et al., 2006; Gehring et al., 2006; Jullien et al., 2006b). Similarly, the FIS-PRC2 complex may be involved in maintaining differential expression at *FH5* after fertilization (Fitz Gerald et al., 2009). However, since nothing is known about the dynamics of H3K27 methylation in the gametes nor during early seed development, the exact role of H3K27 methylation in imprinting remains elusive.

The FIS-PRC2 complex clearly plays a role in establishing the imprinted state of *PHE1*. FIS-PRC2 activity in the female gametes represses the *PHE1* maternal allele. The repressed state of the maternal allele is maintained after fertilization, while the paternal allele somehow escapes silencing by FIS-PRC2 in the fertilization products (Köhler et al., 2005). How the FIS-PRC2 distinguishes the two parental alleles after fertilization is not known. Parental alleles might be differentially marked by

(A) Imprinting control at the *FWA* and *FIS2* locus. In the central cell (CC), MET1 is thought to be repressed by RBR and MSI1, which should result in a passive reduction of DNA methylation. DME removes DNA methylation marks in the CC. In the egg cell (EC), *DME* is not expressed and the locus remains silent. In the sperm cells, MET1 methylates and silences the imprinted gene. After fertilization, the maternal alleles are expressed in the endosperm but not in the embryo.

(B) Imprinting control at the *MEA* locus. In the CC, DME specifically removes the DNA methylation marks. In the EC and potentially in the CC as well (see text), an unknown imprinting factor renders the maternal *MEA* allele active. For simplicity, the autorepression of the maternal *MEA* allele is not shown. In the sperm cells, the paternal *MEA* locus remains methylated and silent through the action of MET1. After fertilization, the silencing of the paternal MEA allele is reinforced by the action of the FIS-PRC2 complex via H3K27 methylation in the endosperm. It remains unclear how *MEA* expression is controlled in the embryo. The maternal *MEA* allele is expressed, while the paternal *MEA* allele is not detected in some accessions, while it is in others.

(C) Imprinting control of *Mez1* and *Fie1* in maize. The genetic factors controlling imprinted gene expression in maize are not known, but the associated epigenetic marks for active and silent chromatin are well described. In the CC, the DNA methylation marks are removed, but the maternal alleles remain silent. For Mez1, epigenetic marks and expression patterns are not known in gametes but might follow the same model. In the egg cell and the sperm cells, the loci remain methylated and silent. In the endosperm, the active maternal allele is marked by H3 and H4 acetylation and H3K4 methylation. The silent allele is repressed by DNA methylation and H3K27 methylation. Both, *Mez1* and *Fie1* are not expressed in the embryo.

(D) Imprinting control at the *Mee1* locus. The maternal *Mee1* allele is active in the endosperm and the embryo. DNA methylation is probably removed after fertilization only, although *Mee1* is weakly expressed in the CC. How exactly the maternal alleles are activated is not known. CC, central cell; EC, egg cell.

additional histone modifications, as proposed in maize (see below), or by differential DNA methylation (see above) and thus may be distinguishable for the FIS-PRC2 after fertilization.

Additional histone marks might be involved in distinguishing the parental alleles of imprinted genes in maize in addition to differential methylation (Haun and Springer, 2008). Chromatin immunoprecipitation with antibodies against specific histone modifications followed by allele-specific RT-PCR has been used to test the abundance of specific histone marks at the *Mez1*, *Fie1*, and *Nrp1* loci in maize. As expected, H3K27 di- and trimethylation, both repressive marks (Peterson and Laniel, 2004), are enriched at all three paternally silent alleles. On the other hand, acetylation of H3 and H4 and dimethylation of H3K4, marks associated with active chromatin (Peterson and Laniel, 2004), are enriched on the active maternal alleles (Figure 2C). The presence of these antagonistic histone marks is specific to the endosperm, since no such enrichment was observed in leaf tissue. Future technical advances might make it possible to investigate the epigenetic profiles of these loci in the gametes to determine whether these marks were established prior to fertilization or correspond to maintenance marks that act downstream of the primary imprint. Whether these chromatin modifications also mark the active and silent alleles at imprinted loci in *Arabidopsis* is yet to be determined.

VERSATILE PLANT IMPRINTING: HOW FLEXIBLE IS IMPRINTING REGULATION IN PLANTS?

A consequence of genomic imprinting is the strict requirement of both parental alleles for the normal development and physiology in mammals. This is best exemplified by human disorders caused by mutations at imprinted loci (Butler, 2009) but also by developmental abnormalities following cloning by somatic nuclear transfer (Niemann et al., 2008). By contrast, certain species of flowering plants produce maternal embryos in the absence of a paternal contribution (apomixis), suggesting a bypass of genomic imprinting requirements in the embryo and sometimes also in the endosperm (Grossniklaus, 2001; Koltunow and Grossniklaus, 2003). This may be achieved by the absence of functional components regulating maternal silencing, such as the FIS-PRC2 complex, or silencing in sperm cells. Evidence for both has been suggested in the apomictic genus *Hieracium* (Tucker et al., 2003; Rodrigues et al., 2010). However, other epigenetic and reversible alterations may operate in nonobligate apomicts able to reproduce both sexually and asexually.

Genomic imprinting may also act as a barrier against interspecific hybridization (Bushell et al., 2003; Gutierrez-Marcos et al., 2003) by creating an imbalance in the relative dosage of maternal and paternal growth regulators (Dilkes and Comai, 2004). Consistent with this hypothesis, maternal *PHE1* was derepressed in nonviable hybrid seeds resulting from an interspecific cross of two species in the genus *Arabidopsis* (Josefsson et al., 2006). Interestingly, increasing the chromosomal dosage of the maternal *Arabidopsis* parent improves hybrid seed viability, possibly by restoring the appropriate balance of maternal repressors of *PHE1* and other FIS-PRC2 targets. Similarly, the level of imprinted *MEA* expression regulates endosperm responses linked to altered parental dosage by reducing maternal *MEA* expression in response to increased paternal dosage (Erilova et al., 2009). Additional quantitative measurements of imprinting regulators and targets in seeds with different parental ploidies showed alterations of both the relative levels and imprinting states of some, but not all, imprinted genes. However, no simple model of parental dosage-dependent regulation of imprinted loci could be drawn from these measurements (Jullien and Berger, 2010; Tiwari et al., 2010). This indicates that more complex mechanisms of imprinting regulation are involved, possibly reflected by the dynamic changes in epigenetic marks observed at imprinted loci in maize.

CONCLUSIONS AND PERSPECTIVES

Genomic imprinting has a major impact on seed development, both by influencing seed growth and viability. The developmental phenotypes of mutants affecting certain imprinted genes in plants are consistent with predictions made by the parental conflict theory, but other theories might also explain the evolution of genomic imprinting, which may have arisen due to distinct selective pressures at different loci. Either alone or in combination, both DNA methylation as well as histone modifications conferred by the FIS-PRC2 complex are involved in imprinting regulation (Choi et al., 2002; Gehring et al., 2006; Jullien et al., 2006a, 2006b). Importantly, DNA methylation is not sufficient to establish imprinted gene expression at all loci described to date. For instance, it is not clear how paternal expression of *HDG3* is achieved, as it is maternally hypomethylated in the endosperm (Gehring et al., 2009). Similarly, DNA methylation does not always correspond to the expression state at imprinted loci in maize, and alleles that show differential methylation in the fertilization products but not in the gametes (Gutiérrez-Marcos et al., 2006; Jahnke and Scholten, 2009) must carry yet unknown primary epigenetic marks. The distinction of primary from secondary marks will be a focus of future research. Furthermore, the complexities of imprinting regulation clearly indicate the existence of additional, yet unknown, factors required for imprinted expression. For instance, the potential involvement of noncodingRNAs, which play an important role in imprinting regulation in mammals (Koerner and Barlow, 2010), has not been rigorously investigated.

Further complexity is added by the fact that *Mee1* (Jahnke and Scholten, 2009), *PHE1* (Köhler et al., 2005), and *MEA* (Vielle-Calzada et al., 1999; Baroux et al., 2006; Spillane et al., 2007; Figure 1) show imprinted or potentially imprinted expression in the embryo. It is not clear how differential activity of parental alleles at these loci is established. While DME-mediated demethylation in the central cell plays a central role for imprinted expression in the endosperm, *DME* is not expressed in the egg cell (Choi et al., 2002). Unlike in the endosperm, where erasure and resetting mechanisms for imprints are not required because it does not contribute to the next generation, such mechanisms must exist for genes with imprinted expression in the plant embryo. Resetting mechanisms ensure that the epigenetic state of the parental alleles is not inherited from one generation to the next. The gender-specific resetting of imprints occurs in the germ line during gametogenesis in mammals (Reik, 2007; Lees-Murdock and Walsh, 2008), but nothing is known about this

process in plants. Although maternal *Mee1* alleles get remethylated during embryogenesis such that both alleles are equally methylated (Jahnke and Scholten, 2009), this does not constitute such a resetting mechanism because it does not lead to a gender-specific distinction of the alleles. Because plants do not have a segregated germ line, the setting of a gender-specific primary imprint can occur only after the lineages for male and female reproductive organs have been separated.

The fact that genomic imprinting in plants is rather versatile and the requirement for a paternal and maternal genome can be bypassed under certain circumstances is important. One fascinating aspect of plant reproduction is the ability of some species to propagate asexually through seeds. Maybe relaxed imprinting requirements were an essential preadaptation for the evolution of apomixis in these taxa. Therefore, apomixis research might benefit from an improved understanding of imprinting regulation and its function in seed development. New technologies allowing the molecular investigation of gametes and improved genome-wide approaches will uncover more imprinted genes and will certainly produce more detailed genome-wide epigenetic maps that add to our understanding of the role and the regulation of genomic imprinting in flowering plants.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Embryo Isolation and RNA Extraction.

ACKNOWLEDGMENTS

We thank Valeria Gagliardini for performing the RT-PCR experiment shown in Figure 1B and our colleagues for stimulating discussions. Our work on imprinting mechanisms is supported by the University of Zürich and grants from the Swiss National Science Foundation (to U.G.) and a fellowship from the Roche Research Foundation (to M.T.R.).

Received November 4, 2010; revised December 29, 2010; accepted January 13, 2011; published January 28, 2011.

REFERENCES

Barlow, D.P. (1993). Methylation and imprinting: from host defense to gene regulation? Science **260:** 309–310.

Barlow, D.P., and Bartolomei, M.S. (2007). Genomic imprinting in mammals. In Epigenetics, C.D. Allis, T. Jenuwein, D. Reinberg, and M.-L. Caparros, eds (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 357–377.

Barlow, D.P., Stöger, R., Herrmann, B.G., Saito, K., and Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the *Tme* locus. Nature **349:** 84–87.

Baroux, C., Gagliardini, V., Page, D.R., and Grossniklaus, U. (2006). Dynamic regulatory interactions of *Polycomb* group genes: *MEDEA* autoregulation is required for imprinted gene expression in *Arabidopsis.* Genes Dev. **20:** 1081–1086.

Baroux, C., Spillane, C., and Grossniklaus, U. (2002). Genomic imprinting during seed development. Adv. Genet. **46:** 165–214.

Barton, S.C., Surani, M.A., and Norris, M.L. (1984). Role of paternal and maternal genomes in mouse development. Nature **311:** 374–376.

Bayer, M., Nawy, T., Giglione, C., Galli, M., Meinnel, T., and Lukowitz, W. (2009). Paternal control of embryonic patterning in *Arabidopsis thaliana.* Science **323:** 1485–1488.

Birchler, J.A. (1993). Dosage analysis of maize endosperm development. Annu. Rev. Genet. **27:** 181–204.

Bushell, C., Spielman, M., and Scott, R.J. (2003). The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. Plant Cell **15:** 1430–1442.

Butler, M.G. (2009). Genomic imprinting disorders in humans: A mini-review. J. Assist. Reprod. Genet. **26:** 477–486.

Cao, X., and Jacobsen, S.E. (2002). Role of the *Arabidopsis DRM* methyltransferases in *de novo* DNA methylation and gene silencing. Curr. Biol. **12:** 1138–1144.

Cattanach, B.M., and Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. Nature **315:** 496–498.

Chaudhuri, S., and Messing, J. (1994). Allele-specific parental imprinting of *dzr1*, a posttranscriptional regulator of zein accumulation. Proc. Natl. Acad. Sci. USA **91:** 4867–4871.

Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S., and Peacock, W.J. (1997). Fertilization-independent seed development in *Arabidopsis thaliana.* Proc. Natl. Acad. Sci. USA **94:** 4223–4228.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis.* Cell **110:** 33–42.

Danilevskaya, O.N., Hermon, P., Hantke, S., Muszynski, M.G., Kollipara, K., and Ananiev, E.V. (2003). Duplicated *fie* genes in maize: Expression pattern and imprinting suggest distinct functions. Plant Cell **15:** 425–438.

DeChiara, T.M., Robertson, E.J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. Cell **64:** 849–859.

Dilkes, B.P., and Comai, L. (2004). A differential dosage hypothesis for parental effects in seed development. Plant Cell **16:** 3174–3180.

Dilkes, B.P., Spielman, M., Weizbauer, R., Watson, B., Burkart-Waco, D., Scott, R.J., and Comai, L. (2008). The maternally expressed WRKY transcription factor *TTG2* controls lethality in interploidy crosses of *Arabidopsis.* PLoS Biol. **6:** 2707–2720.

Dresselhaus, T., Cordts, S., and Lörz, H. (1999). A transcript encoding translation initiation factor eIF-5A is stored in unfertilized egg cells of maize. Plant Mol. Biol. **39:** 1063–1071.

Engel, M.L., Chaboud, A., Dumas, C., and McCormick, S. (2003). Sperm cells of *Zea mays* have a complex complement of mRNAs. Plant J. **34:** 697–707.

Erilova, A., Brownfield, L., Exner, V., Rosa, M., Twell, D., Mittelsten Scheid, O., Hennig, L., and Köhler, C. (2009). Imprinting of the *Polycomb* group gene *MEDEA* serves as a ploidy sensor in *Arabidopsis.* PLoS Genet. **5:** e1000663.

Feil, R., and Berger, F. (2007). Convergent evolution of genomic imprinting in plants and mammals. Trends Genet. **23:** 192–199.

Fitz Gerald, J.N., Hui, P.S., and Berger, F. (2009). *Polycomb* group-dependent imprinting of the actin regulator *AtFH5* regulates morphogenesis in *Arabidopsis thaliana.* Development **136:** 3399–3404.

Garcia, D., Fitz Gerald, J.N., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis.* Plant Cell **17:** 52–60.

Garnier, O., Laoueillé-Duprat, S., and Spillane, C. (2008). Genomic imprinting in plants. Epigenetics **3:** 14–20.

Gehring, M., Bubb, K.L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science **324:** 1447–1451.

Gehring, M., Choi, Y., and Fischer, R.L. (2004). Imprinting and seed development. Plant Cell **16** (suppl.): S203–S213.

Gehring, M., Huh, J.H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes *MEDEA Polycomb* gene self-imprinting by allele-specific demethylation. Cell **124:** 495–506.

Grossniklaus, U. (2001). From sexuality to apomixis: Molecular and genetic approaches. In Advances in Apomixis Research, Y. Savidan, J. Carman, and T. Dresselhaus, eds (Mexico City: CIMMYT Press), pp. 168–211.

Grossniklaus, U. (2009). Plant science: Paternal patterning cue. Science **323:** 1439–1440.

Grossniklaus, U., and Vielle-Calzada, J.P. (1998). Parental conflict and infanticide during embryogenesis. Trends Plant Sci. **3:** 328.

Grossniklaus, U., Vielle-Calzada, J.P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal control of embryogenesis by *MEDEA*, a *Polycomb* group gene in *Arabidopsis.* Science **280:** 446–450.

Guo, M., Rupe, M.A., Danilevskaya, O.N., Yang, X., and Hu, Z. (2003). Genome-wide mRNA profiling reveals heterochronic allelic variation and a new imprinted gene in hybrid maize endosperm. Plant J. **36:** 30–44.

Gutiérrez-Marcos, J.F., Costa, L.M., Biderre-Petit, C., Khbaya, B., O'Sullivan, D.M., Wormald, M., Perez, P., and Dickinson, H.G. (2004). *maternally expressed gene1* is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. Plant Cell **16:** 1288–1301.

Gutiérrez-Marcos, J.F., Costa, L.M., Dal Prà, M., Scholten, S., Kranz, E., Perez, P., and Dickinson, H.G. (2006). Epigenetic asymmetry of imprinted genes in plant gametes. Nat. Genet. **38:** 876–878.

Gutierrez-Marcos, J.F., Pennington, P.D., Costa, L.M., and Dickinson, H.G. (2003). Imprinting in the endosperm: a possible role in preventing wide hybridization. Philos. Trans. R. Soc. Lond. B Biol. Sci. **358:** 1105–1111.

Haig, D., and Graham, C. (1991). Genomic imprinting and the strange case of the insulin-like growth factor II receptor. Cell **64:** 1045–1046.

Haig, D., and Westoby, M. (1989). Parent-specific gene expression and the triploid endosperm. Am. Nat. **134:** 147–155.

Haun, W.J., Laoueillé-Duprat, S., O'connell, M.J., Spillane, C., Grossniklaus, U., Phillips, A.R., Kaeppler, S.M., and Springer, N.M. (2007). Genomic imprinting, methylation and molecular evolution of maize *Enhancer of zeste (Mez)* homologs. Plant J. **49:** 325–337.

Haun, W.J., and Springer, N.M. (2008). Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. Plant J. **56:** 903–912.

Hermon, P., Srilunchang, K.O., Zou, J., Dresselhaus, T., and Danilevskaya, O.N. (2007). Activation of the imprinted *Polycomb* group *Fie1* gene in maize endosperm requires demethylation of the maternal allele. Plant Mol. Biol. **64:** 387–395.

Hsieh, T.-F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D. (2009). Genome-wide demethylation of *Arabidopsis* endosperm. Science **324:** 1451–1454.

Hurst, L.D., and McVean, G.T. (1998). Do we understand the evolution of genomic imprinting? Curr. Opin. Genet. Dev. **8:** 701–708.

Ideraabdullah, F.Y., Vigneau, S., and Bartolomei, M.S. (2008). Genomic imprinting mechanisms in mammals. Mutat. Res. **647:** 77–85.

Ingouff, M., Fitz Gerald, J.N., Guérin, C., Robert, H., Sørensen, M.B., Van Damme, D., Geelen, D., Blanchoin, L., and Berger, F. (2005a). Plant formin *AtFH5* is an evolutionarily conserved actin nucleator involved in cytokinesis. Nat. Cell Biol. **7:** 374–380.

Ingouff, M., Haseloff, J., and Berger, F. (2005b). *Polycomb* group genes control developmental timing of endosperm. Plant J. **42:** 663–674.

Jahnke, S., and Scholten, S. (2009). Epigenetic resetting of a gene imprinted in plant embryos. Curr. Biol. **19:** 1677–1681.

Johnston, A.J., Matveeva, E., Kirioukhova, O., Grossniklaus, U., and Gruissem, W. (2008). A dynamic reciprocal *RBR*-PRC2 regulatory circuit controls *Arabidopsis* gametophyte development. Curr. Biol. **18:** 1680–1686.

Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-dependent loss of gene silencing during interspecies hybridization. Curr. Biol. **16:** 1322–1328.

Jullien, P.E., and Berger, F. (2009). Gamete-specific epigenetic mechanisms shape genomic imprinting. Curr. Opin. Plant Biol. **12:** 637–642.

Jullien, P.E., and Berger, F. (2010). Parental genome dosage imbalance deregulates imprinting in *Arabidopsis.* PLoS Genet. **6:** e1000885.

Jullien, P.E., Katz, A., Oliva, M., Ohad, N., and Berger, F. (2006b). *Polycomb* group complexes self-regulate imprinting of the *Polycomb* group gene *MEDEA* in *Arabidopsis.* Curr. Biol. **16:** 486–492.

Jullien, P.E., Kinoshita, T., Ohad, N., and Berger, F. (2006a). Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. Plant Cell **18:** 1360–1372.

Jullien, P.E., Mosquna, A., Ingouff, M., Sakata, T., Ohad, N., and Berger, F. (2008). Retinoblastoma and its binding partner MSI1 control imprinting in *Arabidopsis.* PLoS Biol. **6:** e194.

Kermicle, J.L. (1970). Dependence of the *R*-mottled aleurone phenotype in maize on mode of sexual transmission. Genetics **66:** 69–85.

Kermicle, J.L., and Alleman, M. (1990). Gametic imprinting in maize in relation to the angiosperm life cycle. Dev. Suppl. 9–14.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. Science **303:** 521–523.

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J.J., Koornneef, M., and Kakutani, T. (2007). Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. Plant J. **49:** 38–45.

Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Imprinting of the *MEDEA Polycomb* gene in the *Arabidopsis* endosperm. Plant Cell **11:** 1945–1952.

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Control of fertilization-independent endosperm development by the *MEDEA Polycomb* gene in *Arabidopsis.* Proc. Natl. Acad. Sci. USA **96:** 4186–4191.

Koerner, M.V., and Barlow, D.P. (2010). Genomic imprinting-an epigenetic gene-regulatory model. Curr. Opin. Genet. Dev. **20:** 164–170.

Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., and Grossniklaus, U. (2003). The *Polycomb* group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1.* Genes Dev. **17:** 1540–1553.

Köhler, C., Page, D.R., Gagliardini, V., and Grossniklaus, U. (2005). The *Arabidopsis thaliana* MEDEA *Polycomb* group protein controls expression of *PHERES1* by parental imprinting. Nat. Genet. **37:** 28–30.

Koltunow, A.M., and Grossniklaus, U. (2003). Apomixis: A developmental perspective. Annu. Rev. Plant Biol. **54:** 547–574.

Lau, M.M., Stewart, C.E., Liu, Z., Bhatt, H., Rotwein, P., and Stewart, C.L. (1994). Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. Genes Dev. **8:** 2953–2963.

Lees-Murdock, D.J., and Walsh, C.P. (2008). DNA methylation reprogramming in the germ line. Adv. Exp. Med. Biol. **626:** 1–15.

Lin, B.-Y. (1984). Ploidy barrier to endosperm development in maize. Genetics **107:** 103–115.

Lippman, Z., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature **430:** 471–476.

Ludwig, S.R., Habera, L.F., Dellaporta, S.L., and Wessler, S.R. (1989). Lc, a member of the maize R gene family responsible for

tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. Proc. Natl. Acad. Sci. **86:** 7092–7096.

Lund, G., Ciceri, P., and Viotti, A. (1995a). Maternal-specific demethylation and expression of specific alleles of zein genes in the endosperm of *Zea mays* L. Plant J. **8:** 571–581.

Lund, G., Messing, J., and Viotti, A. (1995b). Endosperm-specific demethylation and activation of specific alleles of alpha-tubulin genes of *Zea mays* L. Mol. Gen. Genet. **246:** 716–722.

Luo, M., Bilodeau, P., Dennis, E.S., Peacock, W.J., and Chaudhury, A. (2000). Expression and parent-of-origin effects for *FIS2*, *MEA*, and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. Proc. Natl. Acad. Sci. USA **97:** 10637–10642.

Makarevich, G., Villar, C.B.R., Erilova, A., and Köhler, C. (2008). Mechanism of *PHERES1* imprinting in *Arabidopsis.* J. Cell Sci. **121:** 906–912.

McGrath, J., and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell **37:** 179–183.

Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian development: A parental tug-of-war. Trends Genet. **7:** 45–49.

Moore, T., and Mills, W. (2008). Evolutionary theories of imprinting—Enough already! Adv. Exp. Med. Biol. **626:** 116–122.

Niemann, H., Tian, X.C., King, W.A., and Lee, R.S. (2008). Epigenetic reprogramming in embryonic and foetal development upon somatic cell nuclear transfer cloning. Reproduction **135:** 151–163.

Ning, J., Peng, X.-B., Qu, L.-H., Xin, H.-P., Yan, T.-T., and Sun, M. (2006). Differential gene expression in egg cells and zygotes suggests that the transcriptome is restructed before the first zygotic division in tobacco. FEBS Lett. **580:** 1747–1752.

Ohad, N., Margossian, L., Hsu, Y.C., Williams, C., Repetti, P., and Fischer, R.L. (1996). A mutation that allows endosperm development without fertilization. Proc. Natl. Acad. Sci. USA **93:** 5319–5324.

Otto, S.P., and Goldstein, D.B. (1992). Recombination and the evolution of diploidy. Genetics **131:** 745–751.

Peterson, C.L., and Laniel, M.-A. (2004). Histones and histone modifications. Curr. Biol. **14:** R546–R551.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature **447:** 425–432.

Rodrigues, J.C.M., Luo, M., Berger, F., and Koltunow, A.M.G. (2010). *Polycomb* group gene function in sexual and asexual seed development in angiosperms. Sex. Plant Reprod. **23:** 123–133.

Rodrigues, J.C.M., Tucker, M.R., Johnson, S.D., Hrmova, M., and Koltunow, A.M.G. (2008). Sexual and apomictic seed formation in *Hieracium* requires the plant *Polycomb* group gene *FERTILIZATION INDEPENDENT ENDOSPERM.* Plant Cell **20:** 2372–2386.

Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G. (1998). Parent-of-origin effects on seed development in *Arabidopsis thaliana.* Development **125:** 3329–3341.

Solter, D. (1988). Differential imprinting and expression of maternal and paternal genomes. Annu. Rev. Genet. **22:** 127–146.

Spillane, C., Baroux, C., Escobar-Restrepo, J.-M., Page, D.R., Laoueille, S., and Grossniklaus, U. (2004). Transposons and tandem repeats are not involved in the control of genomic imprinting at the *MEDEA* locus in *Arabidopsis.* Cold Spring Harb. Symp. Quant. Biol. **69:** 465–475.

Spillane, C., Schmid, K.J., Laoueillé-Duprat, S., Pien, S., Escobar-Restrepo, J.-M., Baroux, C., Gagliardini, V., Page, D.R., Wolfe, K.H., and Grossniklaus, U. (2007). Positive Darwinian selection at the imprinted *MEDEA* locus in plants. Nature **448:** 349–352.

Spillane, C., Vielle-Calzada, J.-P., and Grossniklaus, U. (2002). Parent-of-origin effects and seed development: Genetics and epigenetics. In The Handbook of Transgenic Food Plants, Y.H. Hui, G.G. Khachatourians, A. McHughen, W.K. Nip, and R. Scorza, eds (New York: Marcel Dekker Inc.), pp. 109–136.

Surani, M.A., Barton, S.C., and Norris, M.L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature **308:** 548–550.

Tiwari, S., Schulz, R., Ikeda, Y., Dytham, L., Bravo, J., Mathers, L., Spielman, M., Guzmán, P., Oakey, R.J., Kinoshita, T., and Scott, R.J. (2008). *MATERNALLY EXPRESSED PAB C-TERMINAL*, a novel imprinted gene in *Arabidopsis*, encodes the conserved C-terminal domain of polyadenylate binding proteins. Plant Cell **20:** 2387–2398.

Tiwari, S., Spielman, M., Schulz, R., Oakey, R.J., Kelsey, G., Salazar, A., Zhang, K., Pennell, R., and Scott, R.J. (2010). Transcriptional profiles underlying parent-of-origin effects in seeds of *Arabidopsis thaliana.* BMC Plant Biol. **10:** 72.

Tucker, M.R., Araujo, A.-C.G., Paech, N.A., Hecht, V., Schmidt, E.D.L., Rossell, J.-B., De Vries, S.C., and Koltunow, A.M.G. (2003). Sexual and apomictic reproduction in *Hieracium* subgenus *pilosella* are closely interrelated developmental pathways. Plant Cell **15:** 1524–1537.

Tycko, B., and Morison, I.M. (2002). Physiological functions of imprinted genes. J. Cell Physiol. **192:** 245–258.

Varmuza, S., and Mann, M. (1994). Genomic imprinting—Defusing the ovarian time bomb. Trends Genet. **10:** 118–123.

Vielle-Calzada, J.P., Baskar, R., and Grossniklaus, U. (2000). Delayed activation of the paternal genome during seed development. Nature **404:** 91–94.

Vielle-Calzada, J.P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M.A., and Grossniklaus, U. (1999). Maintenance of genomic imprinting at the *Arabidopsis MEDEA* locus requires zygotic *DDM1* activity. Genes Dev. **13:** 2971–2982.

Vivian-Smith, A., and Koltunow, A.M. (1999). Genetic analysis of growth-regulator-induced parthenocarpy in *Arabidopsis.* Plant Physiol. **121:** 437–451.

von Wangenheim, K.H., and Peterson, H.P. (2004). Aberrant endosperm development in interploidy crosses reveals a timer of differentiation. Dev. Biol. **270:** 277–289.

Vrana, P.B., Guan, X.J., Ingram, R.S., and Tilghman, S.M. (1998). Genomic imprinting is disrupted in interspecific *Peromyscus* hybrids. Nat. Genet. **20:** 362–365.

Walia, H., Josefsson, C., Dilkes, B., Kirkbride, R., Harada, J., and Comai, L. (2009). Dosage-dependent deregulation of an *AGAMOUS-LIKE* gene cluster contributes to interspecific incompatibility. Curr. Biol. **19:** 1128–1132.

Wilkins, J.F., and Haig, D. (2003). What good is genomic imprinting: the function of parent-specific gene expression. Nat. Rev. Genet. **4:** 359–368.

Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenführer, J., von Mering, C., and Grossniklaus, U. (2010). *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. Curr. Biol. **20:** 506–512.

RECENT PROGRESS IN THE RESEARCH FIELD OF GENOMIC IMPRINTING IN PLANTS

Since the review was published in 2011, there was quite some advancement in the field and I would like to add a few points. In general, genome-wide profiling studies and a few single gene studies expanded the list of imprinted genes tremendously, and novel insights into the function and regulation of imprinted genes have been gained.

Genomics approaches to identify novel imprinted genes

In 2011, a number of research groups performed genome-wide, allele-specific transcriptome profiling studies of hybrid seeds in *Arabidopsis*, maize, and rice to identify genes that are preferentially expressed from one parental allele (Gehring et al., 2011; Hsieh et al., 2011; Luo et al., 2011; McKeown et al., 2011; Pignatta & Gehring, 2012; Waters et al., 2011; Wolff et al., 2011; Zhang et al., 2011). The total number of imprinted genes increased from around 20 (Raissig et al., 2011) to over 300 potentially imprinted plant genes (Gehring et al., 2011; Hsieh et al., 2011; Luo et al., 2011; McKeown et al., 2011; Waters et al., 2011; Wolff et al., 2011; Zhang et al., 2011). However, comparison of the identified candidate MEGs and PEGs revealed little overlap even within one species (Köhler et al., 2012; Pignatta and Gehring, 2012; McKeown et al., 2011). The small reproducibility between different datasets could partly be due to intrinsic, biological features: First, the different studies analyzed different developmental stages and transiently imprinted genes would, therefore, be identified in one dataset but not another. Second, the analysis of different accessions will uncover accession-specific imprinting of genes, which can explain some of the variability between the datasets. In fact, accession-specific imprinting was described in one of those recent studies, where some genes are monoallelically expressed if inherited from one accession but biallelically contributed in the reciprocal cross (Wolff et al., 2011). In addition, technical aspects like sequencing depth, accurate mapping of reads (Degner et al., 2009), quality of available SNP databases, and the purity of the analyzed seed tissue are likely a major source of inconsistency. Wolff and colleagues (2011) analyzed whole seeds and endosperm-specific genes were filtered *in silico* using available transcriptome datasets of seed tissues (Wolff et al., 2011). Obviously, this approach prevents the identification of any genes that are biallelically expressed in sporophytic tissue but imprinted in the fertilization products. Other studies dissected endosperm, embryo, and seed coat but analyzed only late developmental stages, where dissection is applicable (Gehring et al., 2011; Hsieh et al., 2011). Furthermore, manual dissection of seed tissue does not guarantee samples free of sporophytic contamination. Lastly, and probably most important, are statistical aspects of filtering the allele-specific transcriptome for potentially imprinted genes. Actually, the requirement to call a gene parentally biased differed tremendously between the studies and ranged from 90% of all reads that have to derive from one parent (Waters et al., 2011), over 5 times more reads from one parent (Zhang et al., 2011), to simply assessing deviations from the expected 2:1 ratio in the endosperm (Gehring et al., 2011). Interestingly, when the two *Arabidopsis* datasets that analyzed the same accession and almost the same seed stage were filtered with the same strategy, the overlap increased substantially (Gehring et al., 2011). Similarly in

maize, only 50 maize imprinted genes were in common between the 100 and 179 called imprinted genes (Waters et al., 2011; Zhang et al., 2011). 48 of the remaining 129 of the Zhang et al. (2011) dataset had too few reads in the data of Waters and colleagues (2011) to analyze imprinted expression. 15 additional genes would have been called imprinted in the dataset of Waters and colleagues (2011) if the filtering requirement had been relaxed (Pignatta & Gehring, 2012). This suggests that a large part of the difference between the datasets is owed to different statistical pipelines to call imprinted genes, and to different sequencing depths.

Importantly, there are only three novel genes that are imprinted in all three species analyzed, in addition to the previously identified *Polycomb* group genes (Raissig et al., 2011), and, notably, all are PEGs: *YUCCA10*, a flavin monooxygenase (*AT1G48910*, *Os12g08780.1*, *GRMZM2G091819*), *VARIANT IN METHYLATION5* (*VIM5*, *AT1G57800*, *Os04g22240.1*, *AC191534.3*) and *ARID-BRIGHT*, a DNA binding domain protein (*AT4G11400*, *Os10g30944*, *GRMZM2G000404*; Jiang & Köhler, 2012; Pignatta & Gehring, 2012). This suggests that the imprinted status of the three conserved PEGs conferred a selective advantage, and imprinting was maintained even after diversification of monocots and dicots.

Genomics approaches to identify parent-of-origin-dependent allelic expression are also being applied in mammals (Wang et al., 2008; Babak et al., 2008; Gregg et al., 2010). In 2010, an exciting paper described 1300 candidate imprinted loci in the brain of mouse embryos, therefore increasing the number of imprinted genes in mouse by an order of magnitude (Gregg et al., 2010), and the research community was puzzled how so many imprinted candidates could have been missed. However, a recent study reanalyzed the e15 brain dataset published by Gregg et al. (2010) and included their own dataset (e17.5 brain, Deveale et al., 2012). Although similar numbers of candidate imprinted genes were called, the overlap between the two studies was minor. Even more surprisingly, analysis of a dataset that derived from a "mock" reciprocal cross produced nearly as many candidate imprinted genes although there should be none (Deveale et al., 2012). When they applied the false discovery rate (FDR) estimated from the mock reciprocal cross and reanalyzed their "real" reciprocal dataset, they ended up with 42 novel candidates instead of >1000 (Deveale et al., 2012). This study demonstrates how important empirical determination of FDRs is when using next-generation genomic approaches to determine the allele-specific origin of transcripts. In addition, in-depth analysis and substantial validation of imprinted candidates using alternative methods is indispensable.

Different approaches to identify novel imprinted genes

Two independent studies identified novel imprinted genes by different genetic approaches: Shirzadi and colleagues (2011) performed genome-wide transcript profiling of seeds with a normal embryo but an endosperm without paternal contribution, caused by paternally inheriting the *cyclin dependent kinase a;1* (*cdka;1*) mutation (Nowack et al., 2006). Profiling of seeds without paternal contribution identified 600 genes that are downregulated if no paternal genome is inherited in the endosperm (Shirzadi et al., 2011). Type-I MADS-box transcription factors were significantly overrepresented and the authors showed in a series of experiments that *AGAMOUS-LIKE36* (*AGL36*) is only maternally expressed (Shirzadi et al., 2011). Imprinted expression of *AGL36*, that is dispensable for normal seed development, is regulated by DME, MET1 and PRC2, as revealed by mutant crosses and allele-specific analyses (Shirzadi et al.,

2011). Furthermore, Bratzel and colleagues (2012) showed that *AtBMI1C*, one of three PRC1 *BM1* homologues in *Arabidopsis*, is imprinted in the endosperm but shows biallelic expression in the stamen, the male reproductive organ (Bratzel et al., 2012). Its imprinted expression is regulated by differential DNA methylation and the 24nt small interfering RNA (siRNA) pathway since a maternal *dicer-like3* (*dcl3*) mutation and a paternal *met1* mutation derepressed the paternal allele (Bratzel et al., 2012). Interestingly, PRC2 seems to be required for vegetative silencing of *AtBMI1C* but not for its imprinted expression in the endosperm (Bratzel et al., 2012).

Novel regulators of imprinted expression in plants

Recent studies add new regulatory players that aid in controlling imprinted expression in plants. First of all, two of the above mentioned systematic genome-wide hybrid seed transcriptome studies revealed a number of imprinted, long ncRNAs in maize and rice (Luo et al., 2011, Zhang et al., 2011). Interestingly, four of the identified maternally imprinted ncRNAs in maize are transcribed from within four PEGs (Zhang et al., 2011), indicating a similar mode of regulation as in mammals (Santoro and Barlow, 2011). Long ncRNAs have the ability to recruit histone-modifying complexes like PRC2 to potentially silence genes in *cis* (Beisel and Paro, 2011). The mammalian imprinted ncRNAs *Kncq1ot1* and *Airn* are both expressed paternally and recruit histone methyltransferase complexes to silence paternal genes in *cis* (Nagano et al., 2008; Pandey et al., 2008). Also in *Arabidopsis*, the long ncRNA *COLDAIR* recruits PRC2 to induce the vernalization-dependent silencing of *FLOWERING LOCUS C* (*FLC*), a floral repressor, via deposition of H3K27me3 marks (Heo and Sung, 2010). This suggests that long ncRNAs might have a similar role in regulating genomic imprinting in flowering plants as they have in mammals.

In addition, a screen, similar to the one we performed (Chapter 1.2), for activators of expression of the imprinted reporter *FWA::GFP* identified *STRUCTURE SPECIFIC RECOGNITION PROTEIN1* (*SSRP1*), a histone chaperone component, to be required for DNA demethylation in the central cell and, thus, for activation of many parentally imprinted genes (Ikeda et al., 2011). If an *ssrp1* mutation is maternally inherited, then DNA methylation is removed less efficiently from the maternal *FWA* allele. This suggests that maternal SSRP1 is required for active DNA demethylation at the 5' SINE-related repeats of *FWA* and the activation of its maternal allele, although this might be achieved indirectly (Ikeda et al., 2011).

In a yeast two-hybrid screen, Rea and colleagues identified Histone 1 (H1) as a DME-interacting protein and confirmed the interaction by an *in vitro* pull-down assay (Rea et al., 2012). Mutant analysis of *h1* plants revealed that the maternal *H1* allele, but not the paternal allele, is required for DME-mediated regulation of *MEA*, *FWA* and *FIS2* imprinting (Rea et al., 2012). Plants mutant for *H1* show an increase in DNA methylation in the promoter of the maternal allele of *MEA* and *FWA* (Rea et al., 2012). This suggests that H1 might play a role in targeting DME to maternal alleles of imprinted genes, but the exact mechanism of the specific targeting and functional significance remains elusive, since H1 is likely ubiquitously distributed in the genome.

Lastly, a very recent report sequenced the methylome and small RNAs of three haploid cell types from developing or mature pollen: the sperm cells, the vegetative companion cell and their precursor,

the postmeiotic microspore. Among other aspects, they found that TEs are targeted by DME and REPRESSOR OF SILENCING1 (ROS1) for demethylation in the vegetative nucleus, including TEs flanking MEGs (Calarco et al., 2012). In sperm cells, CG methylation is maintained and, in addition, 24nt siRNAs accumulate specifically at transposons flanking the MEGs and might be involved in reinforcing the silent status of the paternal MEG alleles (Calarco et al., 2012) by RNA-dependent DNA methylation (Matzke et al., 2009). The concept of sacrificing genome integrity in companion cells by demethylation and transposon reactivation was proposed several times and for several different tissues: Downregulation of *MET1* during female gametogenesis (Jullien et al., 2008) and activation of *DME* specifically in the central cell (Choi et al., 2002) demethylates the genome of the central cell and, subsequently, the genome of the endosperm globally (Gehring et al., 2009; Hsieh et al., 2009; Ibarra et al., 2012). This likely activates TEs, which then activate the Polymerase IV (PolIV)/PolV-dependent 24nt siRNA machinery, explaining the accumulation of a huge maternal population of 24nt siRNA in the endosperm (Mosher et al., 2009). It is speculated that those 24nt siRNA, if mobile, potentially enforce transposon silencing in the egg cell and, subsequently, the embryo by directing non-CG methylation (Bourc'his and Voinnet, 2010). Similarly in pollen, the accessory vegetative cell activates TEs through the downregulation of *DECREASED DNA METHYLATION1* (*DDM1*, Slotkin et al., 2009) and expression of *DME* (Ibarra et al., 2012; Schoft et al., 2011). As a consequence, the vegetative cell produces an abundance of siRNAs that could potentially enforce transposon silencing in the sperm cells (Slotkin et al., 2009, Ibarra et al., 2012). Therefore, sacrificing genome integrity in tissue that does not contribute to the next generation might enforce genome integrity via protective and mobile siRNAs in the egg cell, the sperm cells, and the embryo (Bourc'his & Voinnet, 2010; Ibarra et al., 2012). Furthermore, it might be aiding in imprinting regulation by reinforcing silencing of TEs close to paternal alleles of MEGs (Calarco et al., 2012).

The maize imprinted gene *Meg1* is involved in nutrient transfer

A recent study in maize highlights the role of the maize imprinted gene *Maternally expressed gene1* (*Meg1,* Gutiérrez-Marcos et al., 2004) in nutrient transfer from mother to offspring. *Meg1* is necessary and sufficient for the differentiation of nutrient transfer cells located at the mother-offspring interface in the maize seed (Costa et al., 2012). In addition, *Meg1* regulates nutrient translocation in the seed and dosage regulation of *Meg1* by genomic imprinting is critical to balance nutrient distribution (Costa et al., 2012). However, unlike predicted by the parental conflict theory (Haig and Westoby, 1989), *Meg1* is a maternal gene that rather promotes than restricts nutrient transfer to the offspring. Thus, evolution of imprinted expression at the *Meg1* locus might be an example of coadaptive evolution of mother and offspring, which predicts that traits important for the mother-offspring interface have been evolved to be under maternal control (Wolf and Hager, 2006).

References

Babak, T., Deveale, B., Armour, C., Raymond, C., Cleary, M.A., van der Kooy, D., Johnson, J.M., and Lim, L.P. (2008). Global survey of genomic imprinting by transcriptome sequencing. Curr Biol **18**: 1735–1741.

Beisel, C. and Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. Nat Rev Genet **12**: 123–135.

Bourc'his, D. and Voinnet, O. (2010). A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. Science **330**: 617–622.

Bratzel, F., Yang, C., Angelova, A., López-Torrejón, G., Koch, M., del Pozo, J.C., and Calonje, M. (2012). Regulation of the new *Arabidopsis* imprinted gene *AtBMI1C* requires the interplay of different epigenetic mechanisms. Mol Plant **5**: 260–269.

Calarco, J.P., Borges, F., Donoghue, M.T.A., Ex, F.V., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D., and Martienssen, R.A. (2012). Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. Cell **151**: 194–205.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. Cell **110**: 33–42.

Costa, L.M., Yuan, J., Rouster, J., Paul, W., Dickinson, H., and Gutierrez-Marcos, J.F. (2012). Maternal control of nutrient allocation in plant seeds by genomic imprinting. Curr Biol **22**: 160–165.

Degner, J.F., Marioni, J.C., Pai, A.A., Pickrell, J.K., Nkadori, E., Gilad, Y., and Pritchard, J.K. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. Bioinformatics **25**: 3207–3212.

Deveale, B., Van Der Kooy, D., and Babak, T. (2012). Critical Evaluation of Imprinted Gene Expression by RNA–Seq: A New Perspective. PLoS Genet **8**: e1002600.

Gehring, M., Bubb, K.L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science **324**: 1447–1451.

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic Analysis of Parent-of-Origin Allelic Expression in *Arabidopsis thaliana* Seeds. PLoS ONE **6**: e23687.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science **329**: 643–648.

Gutiérrez-Marcos, J.F., Costa, L.M., Biderre-Petit, C., Khbaya, B., O'Sullivan, D.M., Wormald, M., Perez, P., and Dickinson, H.G. (2004). *maternally expressed gene1* Is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. Plant Cell **16**: 1288–1301.

Haig, D. and Westoby, M. (1989). Parent-Specific Gene Expression and the triploid Endosperm. American Naturalist **134**: 147–155.

Heo, J.B. and Sung, S. (2010). Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. Science **331**: 76-79.

Hsieh, T.F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D. (2009). Genome-wide demethylation of *Arabidopsis* endosperm. Science **324**: 1451–1454.

Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., and Fischer, R.L. (2011). Inaugural Article: Regulation of imprinted gene expression in *Arabidopsis* endosperm. Proc Natl Acad Sci USA **108**: 1755–1762.

Ibarra, C., Feng, X., Schoft, V., Hsieh, T., Uzawa, R., Rodrigues, J., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R., Tamaru, H., and Zilberman, D. (2012). Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. Science **337**: 1360–1364.

Ikeda, Y., Kinoshita, Y., Susaki, D., Iwano, M., Takayama, S., Higashiyama, T., Kakutani, T., and Kinoshita, T. (2011). HMG domain containing SSRP1 is required for DNA demethylation and genomic imprinting in *Arabidopsis*. Dev Cell **21**: 589–596.

Jiang, H. and Köhler, C. (2012). Evolution, function, and regulation of genomic imprinting in plant seed development. J Exp Bot **63**: 4713–4722.

Jullien, P., Mosquna, A., Ingouff, M., Sakata, T., Ohad, N., Berger, F., and Scott, R. (2008). Retinoblastoma and Its Binding Partner MSI1 Control Imprinting in *Arabidopsis*. PLoS Biol **6**: e194.

Köhler, C., Wolff, P., and Spillane, C. (2012). Epigenetic mechanisms underlying genomic imprinting in plants. Annu Rev Plant Biol **63**: 331–352.

Luo, M., Taylor, J., Spriggs, A., Zhang, H., Wu, X., Russell, S., Singh, M., and Koltunow, A. (2011). A Genome-Wide Survey of Imprinted Genes in Rice Seeds Reveals Imprinting Primarily Occurs in the Endosperm. PLoS Genet **7**: e1002125.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009). RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol **21**: 367–376.

McKeown, P.C., Laouielle-Duprat, S., Prins, P., Wolff, P., Schmid, M.W., Donoghue, M.T., Fort, A., Duszynska, D., Comte, A., Lao, N.T., Wennblom, T.J., Smant, G., Köhler, C., Grossniklaus, U., and Spillane, C. (2011). Identification of imprinted genes subject to parent-of-origin specific expression in *Arabidopsis thaliana* seeds. BMC Plant Biol **11**: 113.

Mosher, R.A., Melnyk, C.W., Kelly, K.A., Dunn, R.M., Studholme, D.J., and Baulcombe, D.C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*. Nature **460**: 283-286.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science **322**: 1717–1720.

Nowack, M.K., Grini, P.E., Jakoby, M.J., Lafos, M., Koncz, C., and Schnittger, A. (2006). A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis. Nat Genet **38**: 63–67.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell **32**: 232–246.

Pignatta, D. and Gehring, M. (2012). Imprinting meets genomics: new insights and new challenges. Curr Opin Plant Biol **15**: 530-535

Raissig, M.T., Baroux, C., and Grossniklaus, U. (2011). Regulation and Flexibility of Genomic Imprinting during Seed Development. Plant Cell **23**: 16–26.

Rea, M., Zheng, W., Chen, M., Braud, C., Bhangu, D., Rognan, T.N., and Xiao, W. (2012). Histone H1 affects gene imprinting and DNA methylation in *Arabidopsis*. Plant J **71**: 776-786.

Santoro, F. and Barlow, D.P. (2011). Developmental control of imprinted expression by macro non-coding RNAs. Semin Cell Dev Biol **22**: 328–335.

Schoft, V.K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J.S., Park, G.T., Fischer, R.L., and Tamaru, H. (2011). Function of the DEMETER DNA glycosylase in the *Arabidopsis thaliana* male gametophyte. Proc Natl Acad Sci USA **108**: 8042-8047.

Shirzadi, R., Andersen, E.D., Bjerkan, K., Gloeckle, B., Heese, M., Ungru, A., Winge, P., Koncz, C., Aalen, R., Schnittger, A., and Grini, P. (2011). Genome-Wide Transcript Profiling of Endosperm without Paternal Contribution Identifies Parent-of-Origin–Dependent Regulation of *AGAMOUS-LIKE36*. PLoS Genet **7**: e1001303.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J.D., Feijó, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136**: 461–472.

Wang, X., Sun, Q., McGrath, S.D., Mardis, E.R., Soloway, P.D., and Clark, A.G. (2008). Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. PloS ONE **3**: e3839.

Waters, A.J., Makarevitch, I., Eichten, S.R., Swanson-Wagner, R.A., Yeh, C.T., Xu, W., Schnable, P.S., Vaughn, M.W., Gehring, M., and Springer, N.M. (2011). Parent-of-Origin Effects on Gene Expression and DNA Methylation in the Maize Endosperm. Plant Cell **23**: 4221–4233.

Wolf, J.B. and Hager, R. (2006). A maternal-offspring coadaptation theory for the evolution of genomic imprinting. PLoS Biol **4**: e380.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C. (2011). High-Resolution Analysis of Parent-of-Origin Allelic Expression in the *Arabidopsis* Endosperm. PLoS Genet **7**: e1002126.

Zhang, M., Zhao, H., Xie, S., Chen, J., Xu, Y., Wang, K., Guan, H., Hu, X., Jiao, Y., Song, W., and Lai, J. (2011). Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. Proc Natl Acad Sci USA **108**: 20042-7

MATERIAL & METHODS

INDEX

NOTE

All material and methods are described within the chapters or the appendices of this thesis. All chapters are written in the form of a publication, containing introduction, material and methods, results, discussion and references. Therefore, I put here an index of all material and methods used during the thesis. I have not used the methods in grey font personally but they were rather applied by co-authors.

PART 1

WHAT TURNS *MEDEA* ON?

REGULATION OF IMPRINTED *MEDEA* EXPRESSION IN THE *ARABIDOPSIS* SEED

PART 1
ABSTRACT

Prior to this thesis, Wöhrmann and colleagues identified the minimal promoter element that is sufficient to confer imprinted and tissue-specific *MEA* expression and postulated that novel, yet unknown imprinting regulators bind to the *MEA*-ICR (see Chapter 1.1; Wöhrmann et al., 2012). To uncover novel regulators, maternal activators and/or paternal repressors, we applied a biochemical and a forward genetic screen. We tried to identify *MEA*-ICR-binding proteins biochemically by performing an electrophoretic mobility shift assay (EMSA, see Chapter 1.2). In addition, plants carrying the *MEA*-ICR reporter gene, *250pMEA::GUS*, were chemically mutagenized using ethane methyl sulfonate (EMS) and screened for ectopic GUS or loss of GUS expression (see Chapter 1.2). Finally, to map candidate mutations regulating imprinted *MEA* expression, which are likely gametophytic-lethal or homozygous-lethal, by next-generation sequencing, we developed SNP Ratio Mapping (SRM). SRM is based on the distinct segregation ratio between linked an unlinked SNPs in the F1 generation of the 2nd backcross (see Chapter 1.3; Lindner*, Raissig* et al. 2012).

Importantly, each subchapter is built like a publication (even if it is not published or submitted like Chapter 1.2) containing introduction, material and methods, results, discussion and separate references.

PART 1
CHAPTER 1

IDENTIFICATION OF A DNA METHYLATION-INDEPENDENT IMPRINTING CONTROL REGION AT THE *ARABIDOPSIS MEDEA* LOCUS

NOTE

All of CHAPTER 1.1 is published as Wöhrmann HJP, Gagliardini V, Raissig MT et al. (2012) *Genes Dev* **26:** 1837-1850.

MTR contributed to this work as follows: After rejection of the manuscript in *Cell* and *Developmental Cell*, MTR integrated the reviewer's comments into the text, rewrote parts of it, added and commented recently published articles related to this work, and reformatted the manuscript for submission to *Genes and Development*. Furthermore, MTR performed sequence element analysis and contributed the paragraph "*MEA*-ICR sequence elements are found up- or downstream of other imprinted loci" including Supplemental Table S3 and S4. In addition, he isolated embryos for bisulfite sequencing for Figure 5 and integrated additional data into Figure 5. Supplemental information to this publication can be found in Appendix A3.

Identification of a DNA methylation-independent imprinting control region at the *Arabidopsis MEDEA* locus

Heike J.P. Wöhrmann,[1] Valeria Gagliardini,[1] Michael T. Raissig,[1] Wendelin Wehrle,[1] Julia Arand,[2] Anja Schmidt,[1] Sascha Tierling,[2] Damian R. Page,[1] Hanspeter Schöb,[1] Jörn Walter,[2] and Ueli Grossniklaus[1,3]

[1]Institute of Plant Biology and Zürich-Basel Plant Science Center, University of Zürich, CH-8008 Zürich, Switzerland; [2]Laboratory of EpiGenetics, Saarland University, D-66041 Saarbrücken, Germany

Genomic imprinting is exclusive to mammals and seed plants and refers to parent-of-origin-dependent, differential transcription. As previously shown in mammals, studies in *Arabidopsis* have implicated DNA methylation as an important hallmark of imprinting. The current model suggests that maternally expressed imprinted genes, such as *MEDEA* (*MEA*), are activated by the DNA glycosylase DEMETER (DME), which removes DNA methylation established by the DNA methyltransferase MET1. We report the systematic functional dissection of the *MEA* *cis*-regulatory region, resulting in the identification of a 200-bp fragment that is necessary and sufficient to mediate *MEA* activation and imprinted expression, thus containing the imprinting control region (ICR). Notably, imprinted *MEA* expression mediated by this ICR is independent of DME and MET1, consistent with the lack of any significant DNA methylation in this region. This is the first example of an ICR without differential DNA methylation, suggesting that factors other than DME and MET1 are required for imprinting at the *MEA* locus.

[*Keywords*: *Arabidopsis*; *DEMETER*; DNA methylation; genomic imprinting; *MEDEA*; imprinting control region]

Supplemental material is available for this article.

Received April 27, 2012; revised version accepted June 25, 2012.

Genomic imprinting is a form of epigenetic gene regulation, which leads to the differential expression of an allele according to its parent of origin. Its discovery dates back to 1970, when Kermicle (1970) described the maternal effect of the *R* gene, which controls maize kernel coloration. Later, an analogous phenomenon was identified in mice when pronuclear transplantation experiments revealed that both maternal and paternal genomes were required to achieve normal development (McGrath and Solter 1984; Surani et al. 1984). Imprinted genes encode for diverse proteins that function in growth and cellular proliferation, typically in extraembryonic tissues involved in nourishing the newly developing organism; i.e., the placenta in mammals and the endosperm in plants (Grossniklaus 2005; Feil and Berger 2007). The endosperm results from double fertilization in angiosperms: While one sperm cell fertilizes the egg cell, giving rise to the embryo, the second sperm cell fuses with the central cell, leading to the development of the endosperm (Maheshwari 1950). Genomic imprinting in mammals and seed plants evolved independently, but likely in response to similar selective pressures that maintain a fine balance between competing interests of the maternal and paternal genomes in resource allocation (Haig and Westoby 1989; Moore and Reik 1996; Messing and Grossniklaus 1999).

Although some imprinted plant genes are also expressed in the embryo, most show preferential expression in the triploid endosperm, and some of them are essential for seed development (for review, see Raissig et al. 2011). *MEDEA* (*MEA*) and *FERTILIZATION-INDEPENDENT SEED2* (*FIS2*) are maternally expressed genes encoding evolutionary conserved *Polycomb* group (PcG) proteins (Grossniklaus et al. 1998; Luo et al. 1999). Plant PcG proteins form several variants of multiprotein complexes that maintain a silenced state of gene expression over many cell divisions through histone modifications (Pien and Grossniklaus 2007). The MEA–FIE (FERTILIZATION-INDEPENDENT ENDOSPERM) complex, which regulates cell proliferation in the endosperm and embryo, contains the PcG proteins MEA, FIS2, FIE, and MULTICOPY SUPPRESSOR OF IRA 1 (MSI1) (Ohad et al. 1999; Luo et al. 2000; Spillane et al. 2000; Köhler et al. 2003a). Mutations in any of these *FIS*

[3]Corresponding author
Email grossnik@botinst.uzh.ch
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.195123.112.

class genes (*mea*, *fie*, *fis2*, and *msi1*) lead to maternal-effect seed abortion (for review, see Grossniklaus et al. 2001), which, in the case of *MEA* and *FIS2*, is due to their maternal-specific expression (Kinoshita et al. 1999; Vielle-Calzada et al. 1999; Jullien et al. 2006b). To date, *PHERES1* (*PHE1*), which is directly regulated by MEA, represents the only well-studied paternally expressed imprinted gene in plants (Köhler et al. 2003b, 2005). While *MEA* and *FIS2* are required for normal seed development (Grossniklaus et al. 1998; Luo et al. 1999), and *PHE1* plays a role in seed abortion in hybrids (Josefsson et al. 2006), two other maternally expressed genes that were reported to be imprinted, *FLOWERING WAGENINGEN* (*FWA*) and *AGAMOUS-LIKE 36* (*AGL36*), are not essential for seed development (Kinoshita et al. 2004; Shirzadi et al. 2011). Recently, several studies using allele-specific RNA profiling of the seed transcriptome describe many novel candidate imprinted genes in *Arabidopsis* (Gehring et al. 2011; Hsieh et al. 2011; Wolff et al. 2011), rice (Luo et al. 2011), and maize (Waters et al. 2011; Zhang et al. 2011). Yet, little is known concerning their role during seed development or their allele-specific regulation.

In contrast, the molecular mechanism underlying the maternal monoallelic expression of *MEA*, *FIS2*, and *FWA*, which results from genomic imprinting (for review, see Grossniklaus 2005), has been studied in some detail. Imprinting of all three loci results from a combination of maternal allele activation and paternal allele silencing. DNA and histone methylation function as epigenetic marks to distinguish maternal and paternal alleles, with DNA methylation playing a critical role in the regulation of all three loci (Vielle-Calzada et al. 1999; Luo et al. 2000; Kinoshita et al. 2004; Jullien et al. 2006b). The current model for imprinting control of *FIS2* and *FWA* involves repressive DNA methylation of both parental alleles by the maintenance DNA methyltransferase MET1 throughout vegetative development. The silencing of the paternal *MEA* allele, however, depends on repressive histone H3 Lys 27 methylation (H3K27me) mediated by a vegetatively acting PcG complex (Jullien et al. 2006a). During male gametogenesis, paternal allele silencing is maintained by MET1 for *FIS2* and *FWA* but by the PcG protein FIE at the paternal *MEA* allele, since in MET1-deficient pollen, the paternal *MEA* allele is not derepressed (Gehring et al. 2006; Jullien et al. 2006a,b). In contrast, during female gametogenesis, the DNA glycosylase DEMETER (DME) removes maternal DNA methylation at all three loci, which results in expression of the maternal allele in the central cell and, subsequently, during seed development (Choi et al. 2002; Kinoshita et al. 2004; Gehring et al. 2006; Jullien et al. 2006b). This demethylation process also involves a histone chaperone, illustrating the interplay of DNA methylation and chromatin level regulation (Ikeda et al. 2011).

In addition to the shared regulation of imprinting at the *FIS2* and *FWA* loci, additional mechanisms appear to operate at the *MEA* locus: *MEA* is expressed in both the embryo and endosperm, and paternal *MEA* allele expression has not been detected during early seed development, suggesting that it is imprinted in both fertilization products at these stages, at least in some accessions (Vielle-Calzada et al. 1999; Luo et al. 2000; Spillane et al. 2007; Raissig et al. 2011). Thus, it is currently unknown how the maternal *MEA* allele is activated in the embryo in the absence of DME activity, which is thought to be restricted to the central cell (Choi et al. 2002). Nevertheless, maternal *MEA* allele activation in the central cell by DME has been the main focus of imprinting regulation in *Arabidopsis*, and possible DME target regions at the *MEA* locus have been identified: The *AtREP2* helitron, CG sites 3 kb and 500 bp upstream of the *MEA* coding region, and the MEA-intergenic subtelomeric repeat (ISR) (Cao and Jacobsen 2002) downstream from the *MEA* coding region were shown to be methylated (Xiao et al. 2003). Indeed, DME establishes allele-specific hypomethylation of the maternal *MEA* allele at the −500-bp region and the MEA-ISR, suggesting that these regions control *MEA*-imprinted expression via their methylation status (Gehring et al. 2006). However, *Arabidopsis* accessions lacking the MEA-ISR remain imprinted at the *MEA* locus (Spillane et al. 2004), and the methylation status of the −500-bp region is not only controlled by DME, but varies depending on the accession; i.e., this region is unmethylated in the Landsberg *erecta* (L*er*) accession, despite *MEA* being imprinted in L*er* (Spillane et al. 2004; Gehring et al. 2006; Schoft et al. 2011). Taken together, this challenges DME as the regulator of imprinted *MEA* expression and raises the question of the actual *cis*-regulatory element for *MEA* imprinting.

Here we report on a minimal 200-base-pair (bp) fragment from the *MEA cis*-regulatory region that faithfully recapitulates *MEA*-like expression and functionally complements the *mea* mutation. Hence, it contains all of the necessary elements for transcriptional activation and imprinting control. We show that activation by *DME* is not mediated by this 200-bp fragment, thereby uncoupling maternal activation by DME from the imprinting control region (ICR). Genetic analysis of seed abortion indicated that DME and MET1 are only indirectly involved in *MEA* imprinting regulation. Maternally, *dme*-induced seed abortion could not be rescued by a functional *MEA* transgene; paternally, rescue of *mea*-induced seed abortion by *met1* mutant pollen was not linked to a functional paternal *MEA* allele. As suggested previously (Gehring et al. 2006), allele-specific expression analysis showed that paternal *MEA* silencing is independent of MET1, consistent with the lack of significant methylation in the *MEA*-ICR. We propose a new model of *MEA* imprinting, in which DME and MET1 affect higher-order chromatin structure through targeting of transposon-related sequences but are not directly involved in the regulation of *MEA* imprinting.

Results

Cis-*activating regions and ICRs reside in the 200-bp* MEA *promoter*

In order to identify the minimal *cis*-regulatory region for imprinted *MEA* expression, we undertook a systematic deletion analysis of the *MEA cis*-regulatory sequences (Fig. 1A). The −4-kb *MEA* upstream sequence, which was

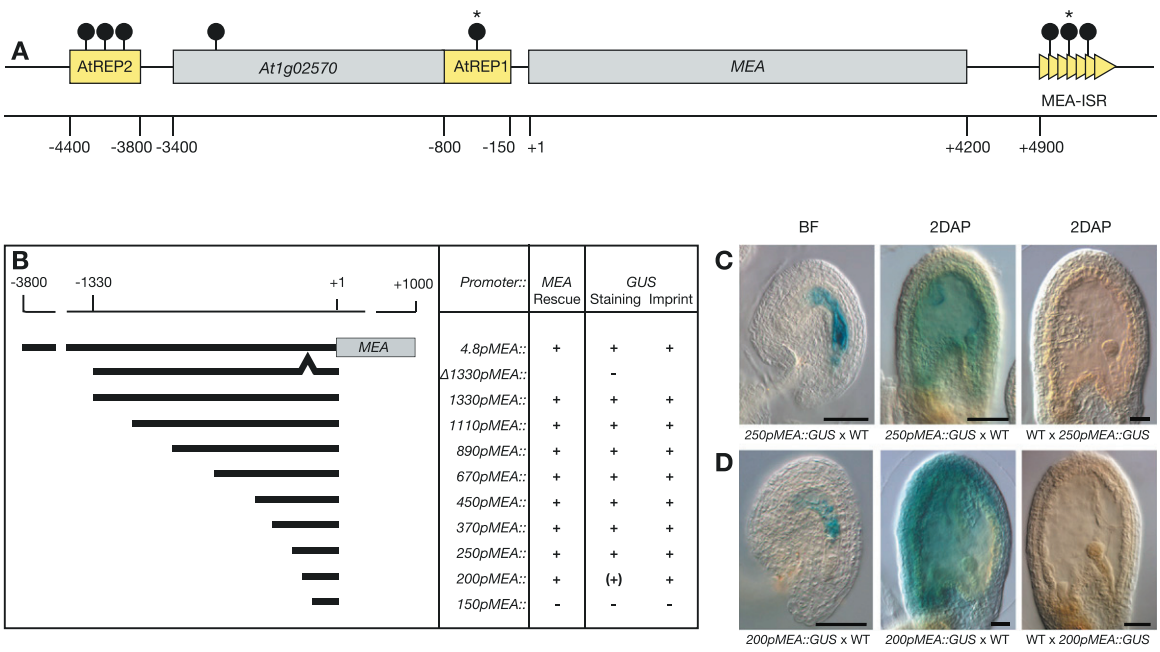


Figure 1. *MEA* promoter dissection. (*A*) The *MEA* locus contains two helitron transposons, *AtREP2* and *AtREP1*, 5′ of the translational start site and a tandem repeat region, termed MEA-ISR, 3′of the gene. *At1g02570* resides in the formerly designated *MEA* promoter (see also Supplemental Fig. S1). Numbers are relative to the translational start site. (Gray boxes) Genes; (yellow boxes) transposons and repeats; (arrowheads) 182-bp direct repeats; (lollipops) sites of DNA methylation as reported (Xiao et al. 2003; Gehring et al. 2006); (stars) hypomethylation of maternal *MEA* endosperm alleles at 7–9 d after pollination (DAP). (*B*) The *4.8pMEA::MEA* transgene contains 3.8 kb of *MEA* upstream sequence fused to *MEA* cDNA and was shown to complement the *mea*-induced seed abortion phenotype (Makarevich et al. 2006). The *4.8pMEA::GUS* transgene was previously described (Spillane et al. 2004) and contains 3.8 kb of *MEA* upstream sequence plus 1 kb of *MEA* coding region. The other transgenes consist of 1330-bp to 150-bp *MEA* promoter sequence fused to *MEA* genomic DNA (*pMEA::MEA*) or the bacterial *uidA* reporter gene (*pMEA::GUS*). In the *Δ1330pMEA::GUS* transgene, the region between the −200-bp and −150-bp *MEA* upstream sequence is deleted. Plus signs [+] indicate positively tested for rescue, staining, or imprinting; minus signs [−] indicate negatively tested for rescue, staining, or imprinting; the plus sign in parenthesis [(+)] indicates deviation from *MEA*-like GUS staining; and empty fields indicate that the corresponding promoter fusion was not tested. (*C*,*D*) Expression of a *250pMEA::GUS* transgene (*C*) and a *200pMEA::GUS* transgene (*D*). The transgenes were reciprocally crossed to L*er* wild-type plants. Maternal GUS activity is detected with both transgenes before fertilization (BF) and 2 DAP. No paternal GUS activity is detected. For detailed GUS expression analysis, see Supplemental Figure S2. Bar, 50 μm.

shown to confer imprinted expression (Spillane et al. 2004), contains the previously unidentified gene *At1g02570*. It was recently annotated based on expressed sequence tags found in transcription profiling studies and encodes a protein of unknown function (Schmid et al. 2003; Castelli et al. 2004). As it resides between regions implicated in *MEA* regulation, we analyzed expression of *At1g02570* by RT–PCR before and after fertilization. We found no expression during early seed development when *MEA* is expressed (Supplemental Fig. S1), suggesting that this gene does not share regulatory *cis*-elements with *MEA*.

Using the previously described *4.8pMEA::GUS* reporter construct, which comprises 3.8 kb of upstream and 1 kb of coding region of the *MEA* gene (Spillane et al. 2004), successive 5′ deletions were introduced, leading to fragment lengths ranging from 1330 bp to 150 bp of *MEA* *cis*-regulatory sequence. We constructed transcriptional fusions to the *Escherichia coli uidA* gene (*pMEA::GUS*), encoding β-glucuronidase (GUS), and to *MEA* genomic DNA (*pMEA::MEA*) for expression and functional analyses, respectively (Fig. 1B). Several independent primary transformants for each transgene were recovered and scored for *MEA*-like expression (Supplemental Tables S1, S2). Only transgenic lines containing a single copy of the insertion, as determined by Southern blot analysis, were chosen for experiments investigating *MEA* imprinting regulation (data not shown).

We studied maternal *GUS* expression from before fertilization until 4 d after pollination (DAP), corresponding to the globular stage of embryo development (Fig. 1C; Supplemental Fig. S2A,C,D). The plant line harboring the *4.8pMEA::GUS* transgene was used as a reference, its GUS-staining pattern reflecting *MEA* expression (Supplemental Fig. S2A). *pMEA::GUS* transgenes with 1330 bp to 250 bp of *MEA* *cis*-regulatory sequences resulted in GUS-staining patterns that were indistinguishable from the *MEA*-like reference pattern: GUS activity was first detected in gynoecia before fertilization, with the entire embryo sac displaying a strong blue staining. After fertilization, GUS activity was found in the embryo, in the free nuclei in the peripheral endosperm, and at the chalazal cyst region of the endosperm. At 4 DAP, weak

GUS activity was detected in globular stage embryos and at the chalazal pole in the endosperm. Therefore, the minimum element necessary to confer *MEA*-like expression resides in the 250-bp *MEA* upstream sequence.

Reducing the fragment length further to only 200 bp of the upstream sequence resulted in a slightly different GUS-staining pattern, which extended into the surrounding sporophytic endothelium (Fig. 1D; Supplemental Fig. S2F,G). The altered expression observed with the *200pMEA::GUS* indicates that a sporophytic repressor-binding site is located between −250 and −200 bp. We observed no GUS activity in plants with the *150pMEA::GUS* transgene and therefore proposed that the 50-bp fragment, which extends from −200 bp to −150 bp, is required for *cis*-activation of *MEA* expression. Indeed, deletion of this 50 bp in the context of the *1330pMEA::GUS* transgene resulted in a loss of expression in all independent primary transformants analyzed (Fig. 1B; Supplemental Table S1). The 50-bp fragment alone did not result in any detectable expression when fused to a *min35S::GUS* transgene (data not shown), indicating that this fragment is necessary but not sufficient for *cis*-activation of *MEA*.

To test for potential loss of imprinting of the reporter transgenes, we reciprocally crossed plants containing the different *pMEA::GUS* transgenes and looked for possible paternal *pMEA::GUS* expression. All reporter transgenes showing *MEA*-like expression were active only when inherited from the mother (Fig. 1C,D; Supplemental Fig. S2A,C,D,F,G), whereas paternally inherited transgenes were silent (Fig. 1C,D; Supplemental Fig. S2B,E,H). Thus, a *cis*-regulatory fragment as short as 200 bp is able to confer imprinted expression to a *GUS* reporter gene, suggesting the presence of an ICR within this fragment.

The 200-bp fragment mediates functional MEA *expression rescuing seed abortion*

In order to functionally test the *pMEA* fragments, we investigated seed abortion in *mea/MEA* plants transformed with *pMEA::MEA* transgenes. Heterozygous *mea/MEA* mutant plants show 50% seed abortion, and all seeds carrying a maternally inherited *mea* mutation abort irrespective of the paternal contribution (Grossniklaus et al. 1998). We scored seed abortion in transgenic *mea/MEA* plants to look for complementation of the *mea*-induced 50% seed abortion phenotype. In all primary transformants except the ones carrying the *150pMEA::MEA* transgene, we found rescue of the *mea* mutant phenotype illustrated by reduced seed abortion frequencies (Supplemental Table S2). Thus, the 200-bp *cis*-regulatory fragment is necessary and sufficient for functional expression of *pMEA::MEA* transgenes, recapitulating the results with the *pMEA::GUS* transgenes at the functional level.

Taken together, our systematic analysis has uncovered a 200-bp minimal fragment of the *MEA cis*-regulatory region that contains the elements necessary and sufficient for transcriptional activation and imprinting control. An additional element between −250 bp and −200 bp is needed to repress sporophytic expression in the ovule. Thus, we used the *250pMEA::GUS* transgene, reflecting *MEA*-like expression, to investigate *MEA* imprinting control in combination with allele-specific expression analyses of the endogenous *MEA* locus.

MEA-*ICR sequence elements are found upstream of or downstream from other potentially imprinted loci*

We investigated whether sequence elements from the *MEA*-ICR were also present at other potentially imprinted loci. To this aim, we performed a WU-BLAST analysis (http://www.arabidopsis.org/wublast/index2.jsp) of the entire *250pMEA* promoter sequence and of the promoter sequence required for proper *MEA*-like expression (100-bp element between −250 and −150 from the *MEA* start codon) against 3 kb of upstream and downstream sequences of all TAIR10 loci (http://www.arabidopsis.org/wublast/index2.jsp). We then compared the output (684 loci) with all potentially imprinted genes that were recently reported (Gehring et al. 2011; Hsieh et al. 2011; McKeown et al. 2011; Wolff et al. 2011). Interestingly, we found that 15 of these recently published imprinted candidate genes do have conserved sequences upstream of or downstream from the respective gene, suggesting that some *MEA*-ICR sequence elements might be conserved between genes regulated by genomic imprinting (see Supplemental Table S3). A permutation test using 1000 randomized gene samples (n = 684) showed that the probability of finding >14 of the recently described imprinted candidate genes by chance is only $P = 0.051$.

In addition, we performed a motif analysis of the *MEA*-ICR and the putative regulatory sequences of the six imprinted candidate genes with the highest similarity scores (i.e., the smallest *P*-values) using the PLACE database (Higo et al. 1999). Interestingly, we found that GT1-binding sites and DOF-binding elements, both of which are abundant in the *MEA*-ICR (nine and five sites, respectively), were also present in the putative regulatory sequences of all six imprinted candidate genes analyzed (Supplemental Table S4). Surprisingly, a pollen-associated binding element, which we speculate might be involved in recruiting repressors to the paternal allele in the male gametophyte, was also found in all of these sequences. An overview of the identified motifs, including other expected *cis*-regulatory elements such as TATABOX5, GATABOX, and a poly-A signal box, is shown in Supplemental Table S4. However, none of these six candidate imprinted genes was analyzed for regulation by *MET1* or *DME*, such that we have no information on their dependence on DNA methylation. Expression of three of the candidates was analyzed in a *fie* mutant background (*At3g19160*, *At2g18880*, and *At4g29650*) (Wolff et al. 2011), but disruption of PRC2 (Polycomb-repressive complex 2) had no effect on their expression.

Taken together, these bioinformatic analyses showed that some sequence elements of the *MEA*-ICR are conserved in putative regulatory sequences of other imprinted loci. Yet these motifs constitute only a small part of the conserved region, as most of the similarity is based on the high A+T content of the *MEA*-ICR (70%). Nevertheless, the imprinted candidate genes with the highest similarity

do share common motifs, such as GT1-binding sites and DOF-binding elements, possibly reflecting conserved regulatory mechanisms.

The MEA-*ICR mediates activation of maternal* MEA *expression independent of DME*

Allele-specific demethylation of the maternal *MEA* allele by DME in the central cell was proposed to selectively activate the maternal *MEA* allele, whereas the paternal *MEA* allele remains silenced (Gehring et al. 2006). However, the *250pMEA::GUS* transgenes are maternally active and paternally silent even though they lack the −500-bp region targeted by DME-dependent demethylation. To elucidate the impact of *DME* on *MEA*-imprinted expression, we analyzed the maternal activity of two *pMEA::GUS* transgenes in the *dme-4* mutant background (Guitton et al. 2004). We crossed plants homozygous for a single locus of either the *4.8pMEA::GUS* or *250pMEA::GUS* transgene to *dme-4/DME* plants and analyzed the progeny for maternal GUS activity. All F1 plants are hemizygous for the *pMEA::GUS* transgene, and half of them are *dme-4/DME* or *DME/DME*, respectively. F1 plants segregating the *dme-4* mutation were emasculated and analyzed for their GUS-staining pattern before fertilization.

In *DME* wild-type plants hemizygous for either *250pMEA::GUS* or *4.8pMEA::GUS*, we observed 50% and 47% GUS staining in unfertilized ovules, respectively, consistent with Mendelian inheritance of the *pMEA::GUS* transgenes by one-half of the female gametophytes (Fig. 2A,B). In plants hemizygous for the *pMEA::GUS* transgene and heterozygous *dme-4/DME*, one-fourth of the ovules are predicted to inherit both the wild-type *DME* allele and the *pMEA::GUS* transgene, whereas one-fourth will inherit the mutant *dme-4* allele along with the *pMEA::GUS* transgene. If *DME* is a direct activator of maternal *MEA* allele expression, we would expect to see only 25% GUS-staining ovules in *dme-4/DME* plants. Indeed, we found a significant reduction ($P = 0.0003$) from 47% to 34% GUS-staining ovules in *dme-4/DME* plants with the *4.8pMEA::GUS* transgene (Fig. 2A,B), suggesting that the *4.8pMEA::GUS* transgene was partly subject to *DME*-dependent repression. In plants hemizygous for *250pMEA::GUS* and *dme-4/DME*, we obtained 46% GUS-staining ovules (Fig. 2A,B). This is not significantly different ($P = 0.9667$) from the 50% GUS staining found in the *DME* wild-type background, suggesting that all of the ovules inheriting the *dme-4* mutation expressed *250pMEA::GUS*.

Thus, *DME* activation of maternal transgene expression before fertilization is dependent on the *MEA* promoter length, which is likely due to the presence of the *AtREP2* helitron 4 kb upstream of *MEA*. This is supported by a previous study, which demonstrated that *4.2pMEA::GUS* and *4.2pMEA::GFP* transgenes containing 450 bp of *AtREP2* are only active when a maternal wild-type *DME* copy is provided (Choi et al. 2002). Our *4.8pMEA::GUS* transgene, containing 3.8 kb of *MEA* upstream sequence with only 100 bp of *AtREP2*, is partially dependent on *DME* activation, whereas maternal activation of the *250pMEA::GUS* transgene is completely independent of *DME* function. As the *250pMEA::GUS* transgene shows exclusive maternal expression, we conclude that DME is not required for imprinting control beyond the native genomic context; i.e., DME is not targeted to the *MEA*-ICR for activation of maternal *MEA* expression.

The MEA-*ICR mediates paternal transgene silencing by maternal MEA*

The *MEA* promoter analysis revealed the existence of a *MEA*-ICR in the 200-bp fragment. Subsequently, we could show that maternal *MEA* allele activation by DME is not targeted to the *MEA*-ICR on the maternal allele. Therefore, we sought to test whether the previously suggested mechanism for paternal *MEA* allele silencing, involving DNA and histone methylation (Gehring et al. 2006; Jullien et al. 2006a,b), is mediated by the *MEA*-ICR.

The MEA–FIE complex represses the *MEA* paternal allele via deposition of repressive H3K27 dimethylation (H3K27me2), which has been found in a region close to the *MEA* transcriptional start site (Gehring et al. 2006). We asked whether the *MEA*-ICR still responds to repression by the MEA–FIE complex. Therefore, we analyzed *GUS* expression in reciprocal crosses of plants homozygous for the *250pMEA::GUS* transgene and homozygous for *mea-1*. Pollination of female plants homozygous for the *250pMEA::GUS* transgene with *mea-1* mutant pollen (Fig. 3A,B) resulted in the same maternal GUS-staining pattern as in females pollinated with wild-type pollen (Supplemental Fig. S2C). Although the *250pMEA::GUS* transgene is imprinted and paternally not expressed after fertilization of a wild-type ovule

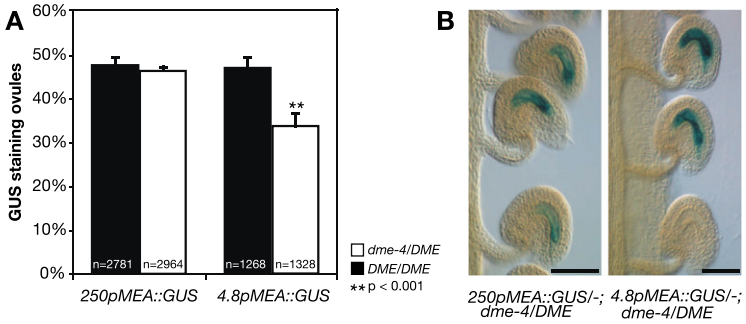


Figure 2. Maternal *MEA* activation by *DME*. (*A*) Percentage of ovules expressing the *250pMEA::GUS* and *4.8pMEA::GUS* reporter transgenes in *DME/DME* and *dme-4/DME* plants before fertilization. At least four independent *DME/DME* and four independent *dme-4/DME* segregants were analyzed for each transgene. Error bars indicate SEM. (n) Total number of ovules analyzed for each genotype; (p) level of significance relative to the difference between the two segregants (*t*-test). (*B*) Maternal *pMEA::GUS* expression of unfertilized ovules in *dme-4/DME* mutant background. Bar, 50 μm.

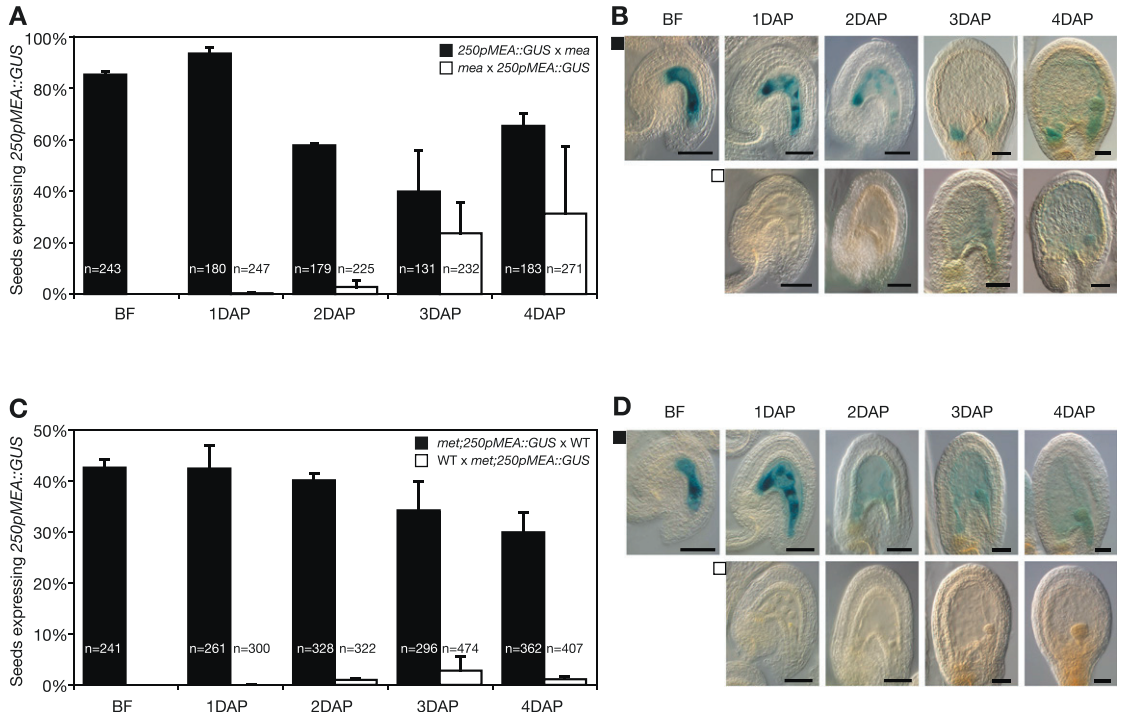


Figure 3. Expression analysis of *250pMEA::GUS* in *mea-1* and *met1-3* mutants. (*A*) Percentage of seeds expressing *250pMEA::GUS* from before fertilization (BF) until 4 DAP. Reciprocal crosses were made between plants homozygous for *250pMEA::GUS* and *mea-1*/*mea-1* (*mea*) plants. Error bars indicate SEM of two biological replicates. (n) Total number of seeds. (*B*) Maternal *250pMEA::GUS* expression was detected throughout seed development (*top* row), and weak paternal *250pMEA::GUS* expression was first detected after 3 DAP (*bottom* row). (*C*) Percentage of seeds expressing *250pMEA::GUS* from before fertilization (BF) until 4 DAP. Reciprocal crosses were made between *met1-3*/*MET* plants hemizygous for *250pMEA::GUS* (*met; 250pMEA::GUS*) and L*er* wild-type plants (WT). Error bars indicate SEM of two biological replicates. (n) Total number of seeds. (*D*) Maternal *250pMEA::GUS* expression was detected throughout seed development (*top* row), and no paternal *250pMEA::GUS* expression was detected (*bottom* row).

(Supplemental Fig. S2E), we found paternal *GUS* expression starting from 3 DAP in maternal *mea-1* mutant plants (Fig. 3A,B). The number of seeds expressing paternal *250pMEA::GUS* in the endosperm increased during development and peaked 4 DAP, with 31% of the seeds showing paternal *250pMEA::GUS* expression. Derepression during 3–4 DAP of the paternally inherited *250pMEA::GUS* transgene in maternal *mea* mutant seeds suggests that the *MEA*-ICR mediates the repressive function of the MEA–FIE complex.

MET1 is not involved in paternal transgene silencing

We found that the maternal MEA protein is required for repression of the paternal *250pMEA::GUS* transgene 3–4 DAP, which is transmitted by the pollen in a transcriptionally silent state (Gehring et al. 2006). As the paternal *MEA* allele provided by *met1* pollen showed no expression in wild-type endosperm 7 DAP, it was concluded that the methylation status of the paternal *MEA* allele is irrelevant for its transcriptional state (Gehring et al. 2006). However, derepression of the paternal *MEA* allele in a *met1* mutant background might only be visible during early seed development, when the MEA–FIE complex is not yet functionally targeting the *MEA* locus.

We tested the impact of *MET1* loss of function during male and female gametogenesis on the expression of the *250pMEA::GUS* transgene. In contrast to previous studies that did not distinguish between indirect effects of *met1-3* due to global DNA hypomethylation (Saze et al. 2003; Mathieu et al. 2007) and direct effects of *met1-3* due to *MEA* hypomethylation, we isolated *met1-3*/*MET* plants from a segregating population of wild-type plants pollinated with *met1-3*/*MET* pollen. In these plants, MET1 activity is missing only in the gametophytes, and thus pre-existing epigenetic misregulation by hypomethylation of genes other than *MEA* can be excluded. We used only *met1-3*/*MET* plants that showed full methylation at the 180-bp centromeric repeat (Martinez-Zapater et al. 1986) as an indication for wild-type methylation levels in those plants. We crossed wild-type pollen to females heterozygous for *met1-3* and hemizygous for *250pMEA::GUS* and investigated maternal GUS activity (Fig. 3C,D). We observed GUS staining in almost all prefertilization ovules and developing seeds after fertilization inheriting the reporter construct (maximum 50%) as in wild-type females (Supplemental Fig. S2C). In contrast, when we used pollen from plants heterozygous for *met1-3* and hemizygous for the *250pMEA::GUS* transgene to fertilize wild-type females, we found no (1 DAP) or only very few (2, 3, and 4 DAP) seeds with paternal

GUS activity (Fig. 3C,D). Thus, a lack of MET1 activity during female or male gametogenesis has no effect on imprinted expression of the *250pMEA::GUS* transgene.

Silencing of the endogenous paternal MEA *allele is controlled by maternal MEA*

The *MEA*-ICR in the paternally inherited *250pMEA::GUS* transgene responds to repression by maternal MEA but not by MET1. In order to correlate the control of paternal transgene silencing with the control of endogenous paternal *MEA* allele silencing, we quantified *MEA* allele-specific transcripts in *mea* mutants and combinations of *mea* mutants with *met1-3* mutants.

We first investigated the role of maternal MEA on paternal *MEA* allele silencing during early seed development. Therefore, we reciprocally crossed *MEA* wild-type plants with *mea* homozygous plants and quantified *MEA* allele-specific transcripts from 1–4 DAP (Fig. 4). Maternal transcripts in reciprocal crosses of *MEA/MEA* and *mea/mea* plants accumulated to their highest level before fertilization and decreased afterward (Fig. 4B). No paternal transcripts were detectable in a maternal *MEA* wild-type background, whereas in a maternal *mea* mutant background, paternal *MEA* allele silencing was released already at 1 DAP (Fig. 4C). Derepression of the paternal *MEA* allele continued until 4 DAP and resulted in more or less constant levels of paternal *MEA* transcripts. Remarkably, the level of derepressed paternal *MEA* transcripts in maternal *mea/mea* mutants represented only 19.5% (0.1563 of 0.8008) of the amount of maternal *MEA* transcripts in the maternal wild-type background (Fig. 4B,C; Supplemental Table S5). However, maternal transcription is no longer autorepressed and highly up-regulated in *mea/mea* mutant plants (Baroux et al. 2006), so the paternal *MEA* transcripts represented only 1.8% (0.1563 of 8.6833) of the amount of maternal *mea* transcripts in *mea/mea* plants (Supplemental Table S5). Thus, derepression of the paternal *MEA* allele in a maternal *mea* mutant background does not result in equivalent expression levels of the two parental alleles. The low level of derepressed paternal *MEA* expression indicates weak paternal *MEA* promoter activity, which might explain why paternal *250pMEA::GUS* expression is only detected 3–4 DAP (Fig. 3A,B). Taken together, we observed derepression of a paternally inherited *250pMEA::GUS* transgene and derepression of the endogenous paternal *MEA* allele in maternal *mea* mutants. This suggests that the *MEA*-ICR is the target of the MEA–FIE complex at the endogenous *MEA* locus.

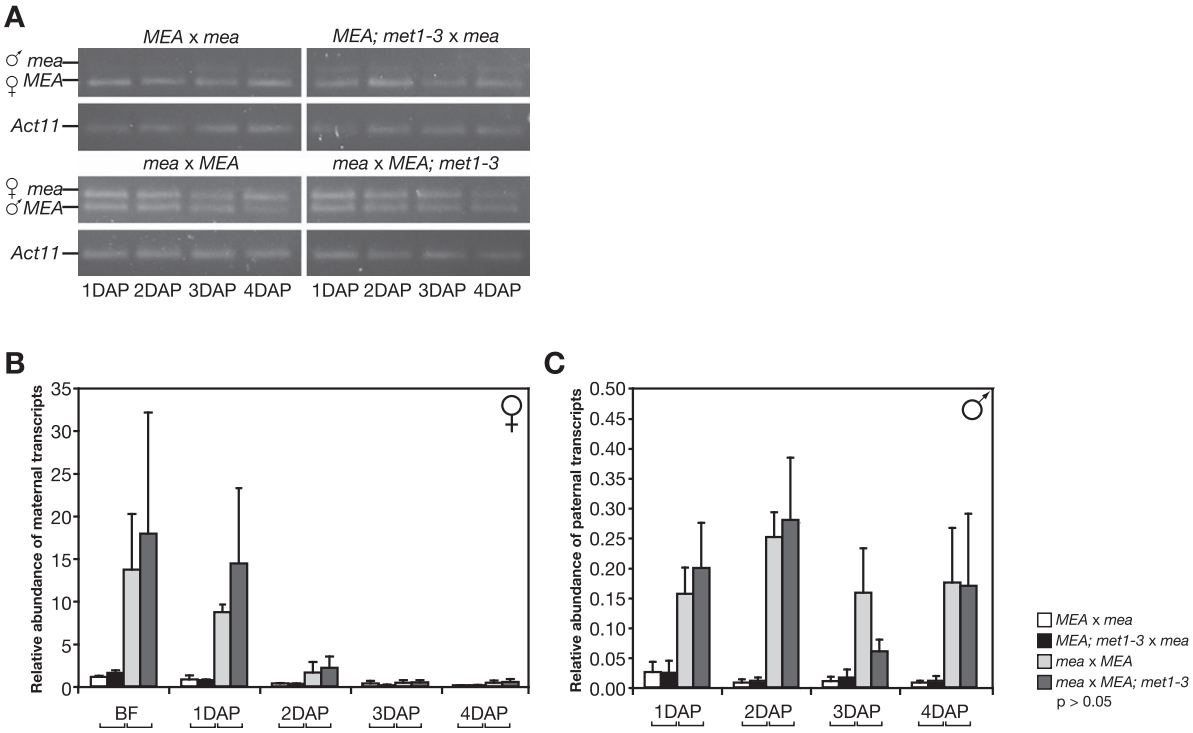


Figure 4. Quantification of *MEA* allele-specific transcription levels. (*A*) Allele-specific RT–PCR on RNA extracted from hand-pollinated siliques at 1–4 DAP. Reciprocal crosses were made between *MEA/MEA* (*MEA*) and *mea-2/mea-2* (*mea*) plants, and between *MEA/MEA; met1-3/MET* (*MEA; met1-3*) and *mea* plants. The RT–PCR products shown are the end products after 40 cycles and show qualitatively whether there is maternal and/or paternal *MEA* expression but are unsuitable to infer quantitative differences. The paternal (♂) and maternal (♀) RT–PCR products are indicated. *Actin 11* (*Act11*) was used as loading control. (*B,C*) Quantification of maternal (*B*) and paternal (*C*) transcripts by RT-qPCR. Transcript levels were normalized to *Act11*. No significant differences in transcript levels were found between crosses with and without *met1-3* (braces *below* the *X*-axis indicate pairwise *t*-tests). Note the different scales for maternal and paternal transcripts. Error bars indicate SEM of three biological replicates.

Since a lack of maternal MEA–FIE PcG activity only leads to very weak derepression of the endogenous paternal *MEA* allele, we wondered whether DNA methylation might be involved in keeping it largely silenced. Thus, we asked again whether MET1 has any residual role in paternal *MEA* allele silencing and crossed *mea*/*mea* mother plants with either wild-type or *met1-3* mutant pollen and analyzed allele-specific *MEA* expression levels (Fig. 4A,C). In *mea*/*mea* mutant mothers, the paternal *MEA* allele was derepressed when transmitted by both wild-type and *met1-3* pollen, with no significant change in the level of derepression (Fig. 4C). This shows that MEA, presumably as part of the maternal MEA–FIE complex, represses the paternal *MEA* allele independent of its methylation status maintained by MET1 during male gametogenesis. Thus, even after removal of both known repressing factors, the maternal MEA–FIE complex and MET1, the paternal *MEA* allele is still expressed at extremely low levels compared with the maternal *MEA* allele; in other words, it is still imprinted. Furthermore, we detected no paternal *MEA* transcripts in the reciprocal cross when *mea* homozygous mutant pollen was crossed to *met1-3* heterozygous females. We conclude that paternal MET1 during male gametogenesis and maternal MET1 during early seed development are not required for *MEA* paternal silencing and thus play no significant role in imprinting at the *MEA* locus.

The MEA-*ICR is unmethylated*

Our comparative analysis of *MEA* transgene and endogene regulation revealed that the *MEA*-ICR is not targeted by DME and MET1. Thus, contrary to what was previously suggested (Gehring et al. 2006; Jullien et al. 2006a), *MEA* imprinting regulation is not primarily controlled by differential DNA methylation. Therefore, we speculated that there is either no DNA methylation at all at the *MEA*-ICR or no differential DNA methylation between active and silent *MEA* alleles.

We analyzed *MEA* promoter methylation in isolated central cells and sperm cells as well as in isolated two-cell stage embryos where the maternal *MEA* allele is expressed (Vielle-Calzada et al. 1999; Spillane et al. 2007). In parallel, we monitored *FWA* promoter methylation, which exhibits imprinting control through a differentially methylated SINE-related element in its promoter (Kinoshita et al. 2004, 2007). In sperm cells, we found high levels of *FWA* promoter methylation in the CG context, consistent with previously reported methylation levels in pollen (Fig. 5E,G; Kinoshita et al. 2004). Surprisingly, we found only a small reduction of CG methylation in the central cell at the *FWA* locus, suggesting that DNA methylation is fully removed after fertilization only (Fig. 5E,F). Contrary to this, we detected almost no methylation in the 250-bp *MEA* promoter from sperm cells and central cells in any sequence context (Fig. 5A–C). In addition, we analyzed methylation in two-cell stage embryos early after fertilization. We detected high methylation levels of *FWA* in the CG contexts in the embryo (Fig. 5E,H), consistent with MET1-dependent silencing of parental *FWA* alleles. However, we found no methylation in the 250-bp *MEA* promoter in the embryo, where the maternal *MEA* allele is expressed (Fig. 5A,D).

In summary, MET1-dependent *FWA* silencing in sperm cells, central cells, and the embryo correlates with DNA methylation in the SINE-related repeat region of its promoter. However, the *MEA*-ICR in the 250-bp *MEA* promoter carries no DNA methylation in any reproductive cell. This confirms our finding that DME is not targeted to the *MEA*-ICR for maternal *MEA* allele activation and that MET1 is not involved in paternal *MEA* allele silencing. Thus, *MEA* is regulated differently from *FIE* and *FWA*, and presently unknown factors, together with the MEA–FIE complex, must be responsible for the imprinted expression of *MEA*.

Discussion

The MEA-*ICR maps to a 200-bp region and displays no differential DNA methylation*

In plants, the primary DNA sequences responsible for genomic imprinting remained elusive. Studies involving transgenes to identify the *cis*-determinants for imprinted expression in *Arabidopsis* and maize indicated that plants ICRs are located close to the imprinted loci (Luo et al. 2000; Kinoshita et al. 2004; Gehring et al. 2006; Gutierrez-Marcos et al. 2006; Makarevich et al. 2008). We identified the 200-bp upstream region adjacent to the *MEA* translational start site as the minimal sequence necessary to confer *cis*-activation and imprinted expression of a *GUS* transgene. The proximity of the *MEA*-ICR and the *MEA* locus is in contrast to mammalian ICRs, which can be located >100 kb distal from the imprinted loci (Ferguson-Smith and Surani 2001).

Mammalian ICRs are typically a few kilobases in length and exhibit parental allele-specific DNA methylation (Bartolomei 2009). However, the *MEA*-ICR maps to a 200-bp fragment and is essentially unmethylated, excluding DNA methylation as the epigenetic mark distinguishing maternal and paternal *MEA* alleles. This is in contrast to the *cis*-elements involved in imprinting at the *FWA* and *PHE1* loci. Maternal-specific expression of *FWA* in the endosperm is due to differential methylation of a SINE-related element located in the *FWA* promoter (Kinoshita et al. 2007). Yet our analysis of DNA methylation in gametes shows that differential methylation at *FWA* is only established after fertilization. This suggests that the primary germline imprint at the *FWA* locus is not the DNA methylation mark itself. Imprinting of *PHE1* results in preferential paternal expression in the endosperm and correlates with differential methylation of tandemly repeated motifs located 3 kb downstream from the *PHE1* gene (Makarevich et al. 2008). Furthermore, differential DNA methylation between the parental alleles has been described for the maize imprinted genes *ZmFie1* and *ZmFie2* (Gutierrez-Marcos et al. 2006). Interestingly, *ZmFie2* is unmethylated in both central cells and sperm cells prior to fertilization, and the differential methylation pattern is only established after fertilization,

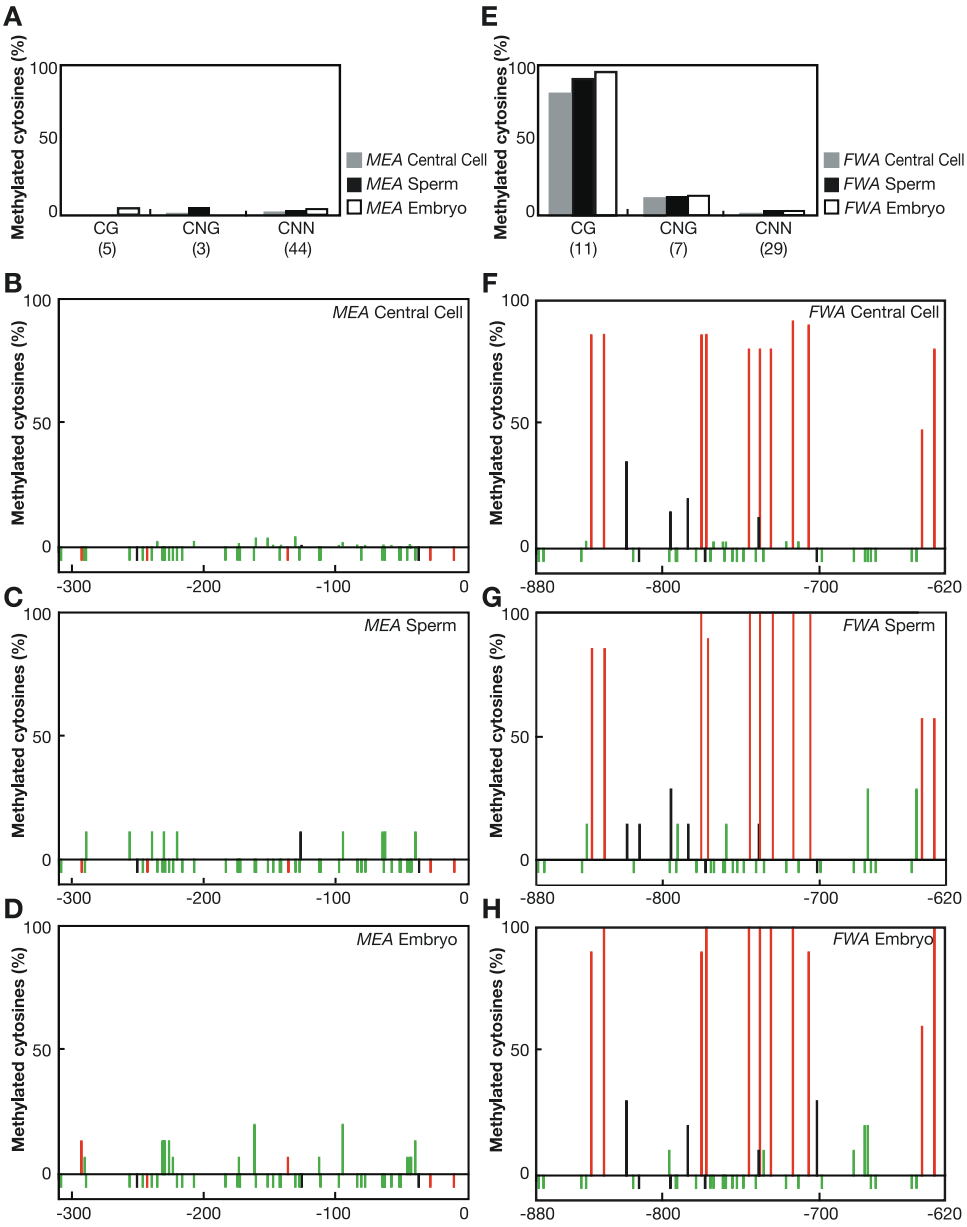


Figure 5. Promoter methylation of *MEA* and *FWA*. (*A*,*E*) Percentage of cytosine methylation at CG, CNG, and CNN sites in the *MEA* promoter (*A*) and the *FWA* promoter (*E*). DNA was isolated from central cells, sperm cells, and two-cell stage embryos; bisulfite-treated; sequenced; and analyzed. Numbers in brackets indicate the number of sites present in the investigated promoter region. (*B–D*,*F–H*) Percentage of cytosine methylation at each position is indicated with a red (CG), black (CNG), or green (CNN) bar. Unmethylated cytosines are shown *below* the 0% line. Numbers are relative to the translational start site and indicate the investigated promoter region.

also indicating that the primary germline imprint is not a DNA methylation mark. In addition, several of the potentially imprinted genes recently identified by transcriptome profiling are unaffected by mutations in one or even all of the known imprinting factors (i.e., *DME*, *MET1*, and *FIE*) (Hsieh et al. 2011; Wolff et al. 2011), suggesting additional, yet-undiscovered imprinting regulators.

MEA is an imprinted gene that is not controlled by differential DNA methylation at the ICR. A related situation may occur in the mouse Prader-Willi/Angelman region showing a complex imprinting control involving several *cis*-acting elements, one of which is not differentially methylated but is required to establish parental imprints at other sites (Kaufman et al. 2009). Moreover, it was recently shown that in macaques, some ICRs that acquire a germline DNA methylation imprint in mice are not methylated in the germline and acquire a differential methylation mark only post-fertilization (A Ferguson-Smith, pers. comm.). Thus, primary imprints that do not involve germline DNA methylation appear to exist in

both plants and mammals. Future studies will show whether common regulatory mechanism indeed exist between nonmethylated ICRs in mammals and plants.

Imprinting control at the MEA-*ICR is independent of DME and MET1*

Maternal allele expression of *MEA* and other maternally expressed imprinted genes depends on the removal of MET1-dependent DNA methylation (Choi et al. 2002; Kinoshita et al. 2004; Jullien et al. 2006b). Consistent with the lack of significant DNA methylation at the *MEA*-ICR, the imprinted *250pMEA::GUS* transgene is maternally activated independent of DME, suggesting that DME is only required in the endogenous context, probably targeting a region different from the *MEA*-ICR. Although involved in imprinting, DNA methylation in flowering plants primarily silences transposons and repeat elements (Henderson and Jacobsen 2007). Thus, a 590-bp *AtREP2* transposon element that is located −4 kb upstream of the *MEA* start codon represents a likely DME target. Indeed, the previously described *4.2pMEA::GUS* transgene containing 450 bp of the *AtREP2* is fully dependent on DME for activation (Choi et al. 2002), whereas the *4.8pMEA::GUS* transgene containing 3.8 kb of *MEA* upstream sequence with 100 bp of the *AtREP2* is only partially dependent on DME (this study). Therefore, we hypothesize that DME is only indirectly involved in the activation of endogenous maternal *MEA* transcription by demethylation of the *AtREP2*.

Based on our results, we propose a new model of *MEA* imprinting regulation (Fig. 6). The methylated *AtREP2* would interact with an unidentified region of the *MEA* locus to establish a silent higher-order chromatin structure; e.g., a repressive chromatin loop. This prevents the *MEA* promoter from being accessed by an unknown transcriptional activator binding the *MEA*-ICR. Demethylation of *AtREP2* by DME in the central cell resolves the repressive chromatin loop and allows the transcriptional activator to access the *MEA*-ICR. The repressive chromatin loop is not resolved in the male gametophyte, where *DME* is not expressed, resulting in exclusive maternal *MEA* allele expression. Since the paternal *MEA* allele is not fully activated if both known repressing activities, MET1 and the MEA–FIE complex, are removed, additional paternal repressors involved in imprinting control have to be postulated, possibly including a PcG complex with a histone methyltransferase other than MEA.

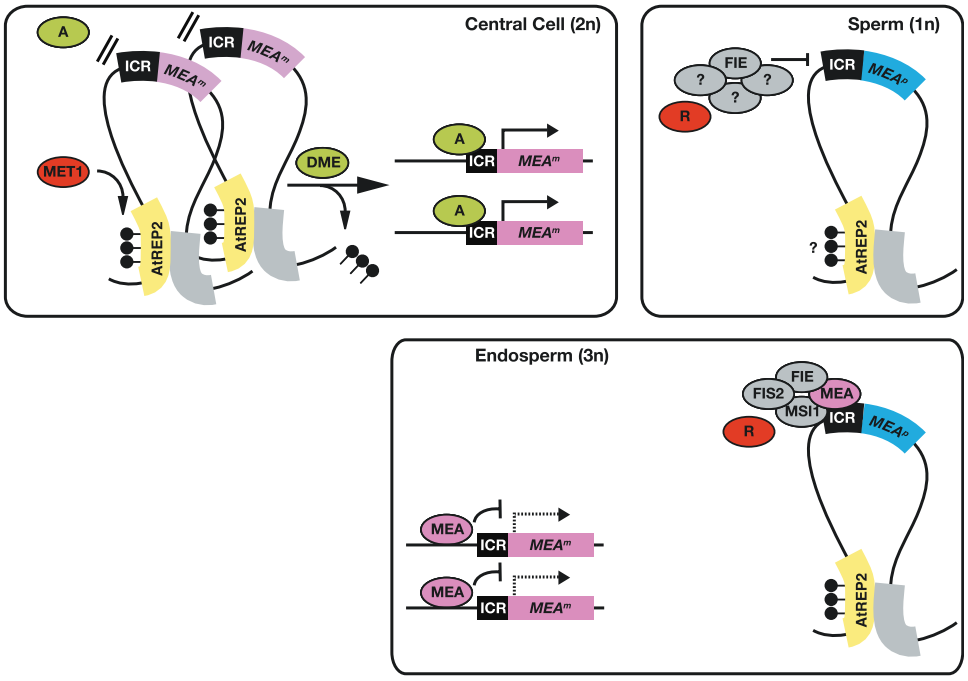


Figure 6. Model of *MEA* imprinting control through a higher-order chromatin structure. Methylation at *AtREP2* is maintained by MET1 in central cells and sperm cells. *AtREP2* might interact with another region, thereby forming a repressive chromatin loop preventing the *MEA* locus from being accessed by a transcriptional activator (A). Specific expression of *DME* in the central cell removes methylation and resolves the repressive chromatin loop. This allows the transcriptional activator (A) to access the *MEA*-ICR. As a consequence, in the endosperm, the two maternal *MEA* alleles (MEA^m) are expressed. Paternal *MEA* allele (MEA^p) silencing is maintained by a proposed PcG complex containing FIE during male gametogenesis (Jullien et al. 2006a). After fertilization, MEA^p is repressed partially by the maternal MEA–FIE complex and another paternal repressor (R). Since parental alleles in the endosperm are differentially targeted by *trans*-acting factors, they must have been marked in the germline, as illustrated by the purple and blue color of the maternal and paternal allele, respectively. The nature of this mark is unknown. The model explicitly shows MEA^m activation in the central cell for imprinted expression in the endosperm; however, the same model is proposed for the egg cell and embryo. (Lollipops) DNA methylation; (dashed line) autorepressed MEA^m transcription.

In mammals, chromosome conformation capture experiments revealed that chromosome looping is involved in imprinting control (Lopes et al. 2003; Kurukuti et al. 2006; Yoon et al. 2007; Engel et al. 2008). More specifically, interactions of differentially methylated regions (DMRs) at the mouse H19/Igf2 locus were shown to partition maternal and paternal chromatin into distinct loops, generating an epigenetic switch to control allele-specific expression (Murrell et al. 2004). Our findings raise the possibility that *MEA* imprinting control might depend on a similar mechanism involving higher-order chromatin structure controlled by DME and MET1.

This hypothesis is consistent with recent reports that *DME* is involved in genome-wide demethylation of the maternal genome in the endosperm, especially of transposons and repeat elements (Gehring et al. 2009; Hsieh et al. 2009). Intriguingly, all characterized imprinted genes in plants are hypomethylated on the maternal allele regardless of which allele is expressed. This suggests that DME-dependent demethylation in the endosperm is not specifically targeting imprinted genes, but rather is a nearly universal process that reshapes DNA methylation of the entire maternal genome in the endosperm.

The imprinting factors required for paternal MEA *silencing remain unknown*

Two epigenetic silencing marks were found at specific sites of the *MEA* locus: DNA methylation and histone H3K27 di- and trimethylation (H3K27me) (Xiao et al. 2003; Gehring et al. 2006; Jullien et al. 2006a). We report that lack of MET1 during male gametogenesis does not derepress the paternal *MEA* allele 1–4 DAP. This complements previous studies with *met1* mutant pollen that showed no paternal *MEA* allele expression 7–9 DAP (Gehring et al. 2006).

Whereas DNA methylation is irrelevant for paternal *MEA* allele silencing, PcG-mediated histone methylation is necessary for paternal *MEA* allele silencing (Gehring et al. 2006; Jullien et al. 2006a). Maternal MEA is involved in deposition of repressive H3K27me at the paternal *MEA* allele close to the translational start site (Gehring et al. 2006). We found derepression of a paternally inherited *250pMEA::GUS* transgene in the maternal *mea* mutant background, suggesting that the *MEA*-ICR in the 250-bp *MEA* promoter is targeted by the maternal MEA–FIE complex. However, it is unclear how the MEA–FIE complex gains access to the silent chromatin loop of the paternal allele to maintain silencing after fertilization. Possibly, the repressive machinery, including the MEA–FIE complex and other proposed repressors, has access to *cis*-regulatory elements in repressive chromatin loops, whereas the activating machinery is efficiently prevented from binding to the *MEA*-ICR.

We found derepression of the paternal *MEA* allele in the maternal *mea* mutant background already at 1 DAP. This contradicts recent findings of delayed paternal derepression, which were explained by the need for passive loss of repressive H3K27me on the paternal *MEA* allele (Jullien et al. 2006a). Surprisingly, derepressed paternal *MEA* transcripts in maternal *mea* mutant plants represent only 14% of maternal *MEA* transcripts in maternal wild-type plants. This resembles the observed residual transcriptional activity of the silent maternal *PHE1* allele (Köhler et al. 2005). Similarly, in mice, paternal alleles of several imprinted genes in the IC2-imprinted domain are not completely silent (Lewis et al. 2004). Even though the silent paternal *MEA* allele is derepressed in *mea* mutant plants, parental transcript levels are clearly not equivalent and still show parent-of-origin-dependent differences. Assuming equivalent parental expression levels in the background of compromised imprinting, the main components involved in paternal *MEA* allele silencing remain to be identified because the paternal *MEA* allele is still imprinted when MET1 and the MEA–FIE complex are missing. As the *MEA*-ICR confers paternal *MEA* silencing beyond the native genomic context, loop formation is not sufficient to explain paternal *MEA* silencing. Thus, another unknown repressor binding to the *MEA*-ICR, along with the proposed PcG complex (Jullien et al. 2006a), may be required for paternal *MEA* repression (Fig. 6).

In summary, our promoter dissection identified the *MEA*-ICR in the 200-bp *MEA* upstream sequence. The *MEA*-ICR carries no significant methylation in sperm cells, central cells, and two-cell stage embryos, which to our knowledge is the first example of an ICR without differential DNA methylation. DME, the key factor necessary for specific activation of maternally expressed imprinted genes in *Arabidopsis*, is dispensable for activation of maternal *MEA* allele transcription. Instead, DME and MET1 may be involved in the regulation of a higher-order chromatin structure at the *MEA* locus, thereby only indirectly controlling the specific marking and activation of the maternal *MEA* allele by unknown factors. However, a repressive chromatin structure at the paternal *MEA* locus alone cannot explain paternal *MEA* silencing, which is mediated through the *MEA*-ICR beyond the native genomic context by still unknown *MEA* imprinting factors.

Material and methods

Plant material

The L*er* accession was used as the wild type. The mutant alleles used were *mea-1*, *mea-2* (L*er*) (Grossniklaus et al. 1998), *dme-4* (C24) (Guitton et al. 2004), and *met1-3* (Col) (Saze et al. 2003). The *4.8pMEA::GUS* transgenic line was described before (Spillane et al. 2004). The *dme-4* (C24) and the *met1-3* (Col-0) mutants were introgressed into the L*er* background by crossing them at least five times as pollen parents. For genotyping assays, methylation status evaluation, and growth conditions, see the Supplemental Material.

Generation of pMEA::MEA *and* pMEA::GUS *constructs*

All *pMEA::MEA* constructs were cloned into pCAMBIA3300 containing the corresponding *MEA* promoter sequence and the entire *MEA* ORF amplified from genomic L*er* DNA. All *pMEA::GUS* constructs contain the corresponding *MEA* promoter sequence amplified from genomic L*er* DNA and were cloned in-frame to the *GUS* reporter gene in pCAMBIA 1381Z.

Promoter deletions were done using different primer pairs amplifying differently sized amplicons and were subsequently cloned in the above-mentioned vectors. For a detailed cloning procedure, see the Supplemental Material.

Microscopy and GUS staining

Histochemical analysis of GUS reporter gene expression was essentially done as described in Baroux et al. 2006. Microscopic inspection was carried out under differential contrast (DIC) optics using a Leica DMR microscope (Leica Microsystems). A detailed description can be found in the Supplemental Material.

RT–PCR analyses

Reverse transcription was performed as previously published (Baroux et al. 2006) on 20 gynoecia before fertilization or on 10–15 siliques at 1–4 DAP, depending on the stage indicated in the corresponding figure. In all experiments, transcript levels were normalized to the level of *ACTIN11* (Huang et al. 1997). For detailed protocol and primers used, see Supplemental Material.

Bisulfite DNA sequencing of isolated reproductive cells

Central cells were isolated using laser capture microscopy, sperm cells were isolated using a Percoll density gradient (M Schauer and U Grossniklaus, unpubl.), and embryos were isolated as previously described (Autran et al. 2011). DNA isolation and bisulfite conversion were essentially performed as described in the Epigenetics Protocols Database "Bisulphite sequencing of small DNA/cell samples" (PROT35; http://www.epigenome-noe.net/research tools/protocols.php). Subsequently, regions of interest (250-bp *MEA* promoter and SINE-related tandem repeat in the *FWA* promoter) were amplified. Purified bisulfite PCR products were cloned into the pGEM-T vector (Promega) and several independent clones were sequenced (for sperm cell and embryo sample), or purified PCR products were directly sequenced with the 454 sequencer according to the standard protocol (central cell samples).

All sequences were analyzed with the BiQ Analyzer software (Bock et al. 2005) for quality control and removal of identical clones in a standardized manner. For a more detailed description, see the Supplemental Material.

Acknowledgments

We thank J. Paszkowski (University of Geneva) and C. Baroux (University of Zürich) for seeds, O. Mittelsten Scheid for the 180-bp repeat probe, M. Schauer for sharing the sperm isolation protocol, K. Kritsas for help with the permutation test, and B. Müller, S. Kessler, and S. Pien for critical comments on the manuscript. This work was supported by the University of Zürich; fellowships of the Forschungskredit der Universität Zürich (to H.J.P.W.), the Roche Research Foundation (to H.S. and M.R.), and the Deutsche Forschungsgemeinschaft (to A.S.); and grants from the Swiss National Science Foundation (to U.G.), Deutsche Forschungsgemeinschaft (to J.W.), and the EPIGENOME Network of Excellence of the European Union (to J.W. and U.G.).

References

Autran D, Baroux C, Raissig MT, Lenormand T, Wittig M, Grob S, Steimer A, Barann M, Klostermeier UC, Leblanc O, et al. 2011. Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. *Cell* **145:** 707–719.

Baroux C, Gagliardini V, Page DR, Grossniklaus U. 2006. Dynamic regulatory interactions of *Polycomb* group genes: *MEDEA* autoregulation is required for imprinted gene expression in *Arabidopsis*. *Genes Dev* **20:** 1081–1086.

Bartolomei MS. 2009. Genomic imprinting: Employing and avoiding epigenetic processes. *Genes Dev* **23:** 2124–2133.

Bock C, Reither S, Mikeska T, Paulsen M, Walter J, Lengauer T. 2005. BiQ Analyzer: Visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* **21:** 4067–4068.

Cao X, Jacobsen SE. 2002. Locus-specific control of asymmetric and CpNpG methylation by the *DRM* and *CMT3* methyltransferase genes. *Proc Natl Acad Sci* **99:** 16491–16498.

Castelli V, Aury JM, Jaillon O, Wincker P, Clepet C, Menard M, Cruaud C, Quétier F, Scarpelli C, Schächter V, et al. 2004. Whole-genome sequence comparisons and 'full-length' cDNA sequences: A combined approach to evaluate and improve *Arabidopsis* genome annotation. *Genome Res* **14:** 406–413.

Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL. 2002. *DEMETER*, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell* **110:** 33–42.

Engel N, Raval AK, Thorvaldsen JL, Bartolomei SM. 2008. Three-dimensional conformation at the H19/Igf2 locus supports a model of enhancer tracking. *Hum Mol Genet* **17:** 3021–3029.

Feil R, Berger F. 2007. Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet* **23:** 192–199.

Ferguson-Smith AC, Surani MA. 2001. Imprinting and the epigenetic asymmetry between parental genomes. *Science* **293:** 1086–1089.

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL. 2006. *DEMETER* DNA glycosylase establishes *MEDEA Polycomb* gene self-imprinting by allele-specific demethylation. *Cell* **124:** 495–506.

Gehring M, Bubb KL, Henikoff S. 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* **324:** 1447–1451.

Gehring M, Missirian V, Henikoff S. 2011. Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS ONE* **6:** e23687. doi: 10.1371/journal.pone.0023687.

Grossniklaus U. 2005 Genomic imprinting in plants: A predominantly maternal affair. In *Annual plant reviews: Plant epigenetics* (ed. P Meyer), pp. 174–200, Blackwell, Sheffield, UK.

Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB. 1998. Maternal control of embryogenesis by *MEDEA*, a *Polycomb* group gene in *Arabidopsis*. *Science* **280:** 446–450.

Grossniklaus U, Spillane C, Page DR, Kohler C. 2001. Genomic imprinting and seed development: Endosperm formation with and without sex. *Curr Opin Plant Biol* **4:** 21–27.

Guitton AE, Page DR, Chambrier P, Lionnet C, Faure JE, Grossniklaus U, Berger F. 2004. Identification of new members of FERTILISATION INDEPENDENT SEED *Polycomb* group pathway involved in the control of seed development in *Arabidopsis thaliana*. *Development* **131:** 2971–2981.

Gutierrez-Marcos JF, Costa LM, Dal Pra M, Scholten S, Kranz E, Perez P, Dickinson HG. 2006. Epigenetic asymmetry of imprinted genes in plant gametes. *Nat Genet* **38:** 876–878.

Haig D, Westoby M. 1989. Parent-specific gene expression and the triploid endosperm. *Am Nat* **134:** 147.

Henderson IR, Jacobsen SE. 2007. Epigenetic inheritance in plants. *Nature* **447:** 418–424.

Higo K, Ugawa Y, Iwamoto M, Korenaga T. 1999. Plant *cis*-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res* **27:** 297–300.

Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D. 2009. Genome-wide demethylation of *Arabidopsis* endosperm. *Science* **324:** 1451–1454.

Hsieh TF, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, et al. 2011. Inaugural article: Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc Natl Acad Sci* **108:** 1755–1762.

Huang S, An YQ, McDowell JM, McKinney EC, Meagher RB. 1997. The *Arabidopsis ACT11* actin gene is strongly expressed in tissues of the emerging inflorescence, pollen, and developing ovules. *Plant Mol Biol* **33:** 125–139.

Ikeda Y, Kinoshita Y, Susaki D, Ikeda Y, Iwano M, Takayama S, Higashiyama T, Kakutani T, Kinoshita T. 2011. HMG domain containing SSRP1 is required for DNA demethylation and genomic imprinting in *Arabidopsis*. *Dev Cell* **21:** 589–596.

Josefsson C, Dilkes B, Comai L. 2006. Parent-dependent loss of gene silencing during interspecies hybridization. *Curr Biol* **16:** 1322–1328.

Jullien PE, Katz A, Oliva M, Ohad N, Berger F. 2006a. *Polycomb* group complexes self-regulate imprinting of the *Polycomb* group gene *MEDEA* in *Arabidopsis*. *Curr Biol* **16:** 486–492.

Jullien PE, Kinoshita T, Ohad N, Berger F. 2006b. Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. *Plant Cell* **18:** 1360–1372.

Kaufman Y, Heled M, Perk J, Razin A, Shemer R. 2009. Protein-binding elements establish in the oocyte the primary imprint of the Prader-Willi/Angelman syndromes domain. *Proc Natl Acad Sci* **106:** 10242–10247.

Kermicle JL. 1970. Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics* **66:** 69–85.

Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL. 1999. Imprinting of the *MEDEA Polycomb* gene in the *Arabidopsis* endosperm. *Plant Cell* **11:** 1945–1952.

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T. 2004. One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303:** 521–523.

Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJ, Koornneef M, Kakutani T. 2007. Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J* **49:** 38–45.

Köhler C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Gruissem W. 2003a. *Arabidopsis MSI1* is a component of the MEA/FIE *Polycomb* group complex and required for seed development. *EMBO J* **22:** 4804–4814.

Köhler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U. 2003b. The *Polycomb* group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. *Genes Dev* **17:** 1540–1553.

Köhler C, Page DR, Gagliardini V, Grossniklaus U. 2005. The *Arabidopsis thaliana* MEDEA *Polycomb* group protein controls expression of *PHERES1* by parental imprinting. *Nat Genet* **37:** 28–30.

Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R. 2006. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci* **103:** 10684–10689.

Lewis A, Mitsuya K, Umlauf D, Smith P, Dean W, Walter J, Higgins M, Feil R, Reik W. 2004. Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nat Genet* **36:** 1291–1295.

Lopes S, Lewis A, Hajkova P, Dean W, Oswald J, Forne T, Murrell A, Constância M, Bartolomei M, Walter J, et al. 2003. Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. *Hum Mol Genet* **12:** 295–305.

Luo M, Bilodeau P, Koltunow A, Dennis ES, Peacock WJ, Chaudhury AM. 1999. Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. *Proc Natl Acad Sci* **96:** 296–301.

Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A. 2000. Expression and parent-of-origin effects for *FIS2, MEA*, and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. *Proc Natl Acad Sci* **97:** 10637–10642.

Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A. 2011. A Genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet* **7:** e1002125. doi: 10.1371/journal.pgen.1002125.

Maheshwari P. 1950. An introduction to the embryology of angiosperms. In *McGraw Hill publications in the botanical sciences* (ed. EW Sinnott), McGraw Hill Book Company, New York.

Makarevich G, Leroy O, Akinci U, Schubert D, Clarenz O, Goodrich J, Grossniklaus U, Köhler C. 2006. Different *Polycomb* group complexes regulate common target genes in *Arabidopsis*. *EMBO Rep* **7:** 947–952.

Makarevich G, Villar CB, Erilova A, Kohler C. 2008. Mechanism of *PHERES1* imprinting in *Arabidopsis*. *J Cell Sci* **121:** 906–912.

Martinez-Zapater JM, Estelle MA, Somerville CR. 1986. A highly repeated DNA sequence in *Arabidopsis thaliana*. *Mol Gen Genet* **204:** 417–423.

Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J. 2007. Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* **130:** 851–862.

McGrath J, Solter D. 1984. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37:** 179–183.

McKeown PC, Laouielle-Duprat S, Prins P, Wolff P, Schmid MW, Donoghue MTA, Fort A, Duszynska A, Comte A, Thi Lao N, et al. 2011. Identification of imprinted genes subject to parent-of-origin specific expression in *Arabidopsis thaliana* seeds. *BMC Plant Biol* **11:** 113.

Messing J, Grossniklaus U. 1999. Genomic imprinting in plants. *Results Probl Cell Differ* **25:** 23–40.

Moore T, Reik W. 1996. Genetic conflict in early development: Parental imprinting in normal and abnormal growth. *Reproduction* **1:** 73–77.

Murrell A, Heeson S, Reik W. 2004. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat Genet* **36:** 889–893.

Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, Goldberg RB, Fischer RL. 1999. Mutations in *FIE*, a WD *Polycomb* group gene, allow endosperm development without fertilization. *Plant Cell* **11:** 407–416.

Pien S, Grossniklaus U. 2007. *Polycomb* group and *trithorax* group proteins in *Arabidopsis*. *Biochim Biophys Acta* **1769:** 375–382.

Raissig MT, Baroux C, Grossniklaus U. 2011. Regulation and flexibility of genomic imprinting during seed development. *Plant Cell* **23:** 16–26.

Saze H, Mittelsten Scheid O, Paszkowski J. 2003. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet* **34:** 65–69.

Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar B. 2003. Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Res* **13:** 1250–1257.

Schoft VK, Chumak N, Choi Y, Hannon M, Garcia-Aguilar M, Machlicova A, Slusarz L, Mosiolek M, Park JS, Park GT, et al. 2011. Function of the *DEMETER* DNA glycosylase in the *Arabidopsis thaliana* male gametophyte. *Proc Natl Acad Sci* **108:** 8024–8047.

Shirzadi R, Andersen ED, Bjerkan KN, Gloeckle BM, Heese M, Ungru A, Winge P, Koncz C, Aalen RB, Schnittger A, et al. 2011. Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of *AGAMOUS-LIKE36*. *PLoS Genet* **7:** e1001303. doi: 10.1371/journal.pgen.1001303.

Spillane C, MacDougall C, Stock C, Kohler C, Vielle-Calzada JP, Nunes SM, Grossniklaus U, Goodrich J. 2000. Interaction of the *Arabidopsis Polycomb* group proteins *FIE* and *MEA* mediates their common phenotypes. *Curr Biol* **10:** 1535–1538.

Spillane C, Baroux C, Escobar-Restrepo JM, Page DR, Laoueille S, Grossniklaus U. 2004. Transposons and tandem repeats are not involved in the control of genomic imprinting at the *MEDEA* locus in *Arabidopsis*. *Cold Spring Harb Symp Quant Biol* **69:** 465–475.

Spillane C, Schmid KJ, Laoueille-Duprat S, Pien S, Escobar-Restrepo JM, Baroux C, Gagliardini V, Page DR, Wolfe KH, Grossniklaus U. 2007. Positive darwinian selection at the imprinted *MEDEA* locus in plants. *Nature* **448:** 349–352.

Surani MA, Barton SC, Norris ML. 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308:** 548–550.

Vielle-Calzada JP, Thomas J, Spillane C, Coluccio A, Hoeppner MA, Grossniklaus U. 1999. Maintenance of genomic imprinting at the *Arabidopsis MEDEA* locus requires zygotic *DDM1* activity. *Genes Dev* **13:** 2971–2982.

Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh C-T, Xu W, Schnable PS, Vaughn MW, Gehring M, Springer NM. 2011. Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* **23:** 4221–4233.

Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MTA, Spillane C, Nordborg M, Rehmsmeier M, Köhler C. 2011. High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet* **7:** e1002126. doi: 10.1371/journal.pgen.1002126.

Xiao W, Gehring M, Choi Y, Margossian L, Pu H, Harada JJ, Goldberg RB, Pennell RI, Fischer RL. 2003. Imprinting of the *MEA Polycomb* gene is controlled by antagonism between *MET1* methyltransferase and *DME* glycosylase. *Dev Cell* **5:** 891–901.

Yoon YS, Jeong S, Rong Q, Park KY, Chung JH, Pfeifer K. 2007. Analysis of the H19ICR insulator. *Mol Cell Biol* **27:** 3499–3510.

Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, Zhao H, Guan H, Hu X, Jiao Y, et al. 2011. Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci* **108:** 20042–20047.

PART 1
CHAPTER 2

BIOCHEMICAL AND GENETIC SCREENS TO IDENTIFY NOVEL REGULATORS OF IMPRINTED *MEDEA* EXPRESSION

INTRODUCTION

To uncover novel regulators of imprinted *MEA* expression in *Arabidopsis thaliana*, we conducted two different screens: First, we tried to identify and isolate proteins binding to the 250 base pair (bp) minimal promoter element (*MEA*-ICR) or parts of it biochemically by applying an electrophoretic mobility shift assay (EMSA). Second, a mutagenesis of plants carrying the *MEA*-ICR reporter gene *250pMEA::GUS* (Wöhrmann et al., 2012; Chapter 1.1) and a forward genetic screen were implemented to find factors regulating imprinted *MEA* reporter gene expression genetically.

EMSA or band shift assays can be applied to show protein-DNA, protein-RNA, or even protein-protein interactions. The technology is based on the different electrophoretic running time of a free probe (i.e. a labeled oligonucleotide) compared to the oligonucleotide bound to a protein, since travelling time through a gel is determined by size and charge. Therefore, the oligonucleotide bound to a protein creates a "band shift" upon polyacrylamide gel electrophoresis (PAGE; Garner and Revzin, 1981). Usually, this technique is used to demonstrate an interaction between a DNA element, usually an enhancer or a promoter element, with an isolated DNA-binding protein. Here, we tried to apply this method to screen a nuclear protein extract of broccoli (*Brassica oleracea* var. *italica*) using radioactively labeled oligonucleotides containing the full *MEA*-ICR or parts of it. We used *B. oleracea* var. *italica* to isolate proteins, because broccoli consists of unopened flower buds, which is the required developmental stage where maternal *MEA* gets activated and starts to be expressed (Grossniklaus et al., 1998). We could not find a positive and specific interaction; therefore, we could not test whether it is possible to cut out the band shift (including the interacting protein) and analyze the interacting partner by mass spectrometry.

To genetically identify factors controlling imprinted *MEA* expression we conducted a forward genetic screen, which is a powerful approach to uncover novel genes and their functions in genetic model organisms like *Arabidopsis thaliana*. We mutagenized plants homozygous for the minimal *MEA-ICR* reporter gene (*250pMEA::GUS*) using the chemical mutagen ethane methyl sulfonate (EMS). To identify novel maternal activator mutants, we screened for absence of maternal reporter expression in the seed and, in a second step, scored for a seed abortion phenotype to reduce false positives (Figure 1-2-2). To uncover paternal repressor candidates, we crossed individuals of the first generation after mutagenesis (M1) as fathers to wild-type mothers and examined the resulting F1 seed for ectopic paternal reporter expression (Figure 1-2-2). We screened approximately 2500 M1 plants after two independent rounds of EMS mutagenesis and found and confirmed (in the F1 of the 2nd backcross (BC2)) two maternal activator candidates and two paternal repressor candidates. SRM was applied to identify the causative SNP of the maternal activator candidates but we were not able to pinpoint the causative mutation due to the genetic background of the *250pMEA::GUS* line: wild-type segregants of a heterozygous and gamma-ray irradiated *demeter* (*dme-4*)/*DME* population were used for transformation with the *250pMEA::GUS* construct (see below). We identified almost a hundred times more SNPs than expected, making the identification of the causative SNP by SRM impossible. Currently, we are crossing the mutation out to Col-0 to create a near isogenic line (NIL) including only a small fragment of the original ecotype that contains the causative mutation.

MATERIAL AND METHODS

EMSA:

Preparation of nuclear proteins:

Young flower buds from *B. oleracea* var. *italica* were collected (6 g) and homogenized, and nuclei were extracted following a published method that is capable of isolating transcriptionally active nuclei (Folta and Kaufman, 2006).

To break up nuclei and produce the crude nuclear protein extract used for EMSA, we centrifuged 100 µl of storage buffer and nuclei suspension for 15 min at 1000g and 4°C, and resuspended the nuclei in the appropriate binding buffer. To break up the nuclei, we sonicated the samples 5 min with 30 sec pulses and 30 sec pause at the highest intensity. This resulted in a protein concentration of ~10 µg/µl, and sonication was as efficient as nuclear breakdown by adding Urea to the nuclei suspension (Figure 1-2-1). Sonication was followed by a centrifugation step (10min at 1000g and 4°C). and binding assays were performed with the resulting supernatant and a total of 20 µg to 80 µg of crude protein per reaction.

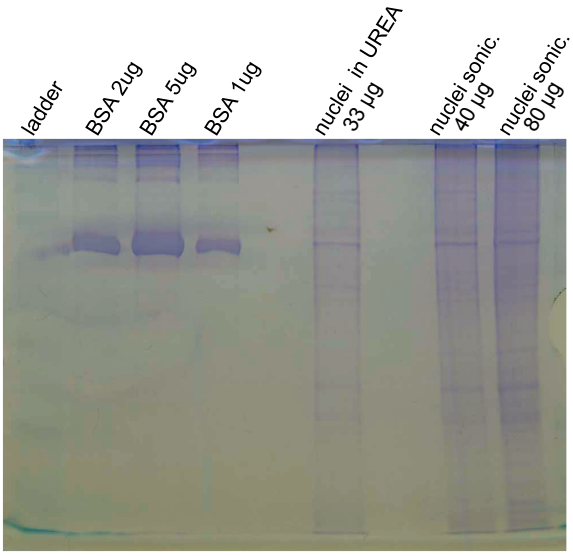


Figure 1-2-1. SDS-PAGE of broccoli nuclear extract proteins. Urea and sonication-based extraction of crude nuclear extracts were compared. Isolated broccoli nuclei were centrifuged and resuspended either in Urea-buffer and incubated at 65°C for 5min or directly in binding buffer and sonicated for 5 min. Different amounts of BSA were loaded as control. SDS-PAGE was run at 20mA for 1h and Coomassie stained for 15min.

Labeling the oligonucleotide probe

We assayed five different oligonucleotides: (i) the full length *MEA*-ICR of 250bp, and (ii) 50bp oligomers covering -250 to -200 bp, -200 to -150 bp, -225 to -175 bp, and -175 to -125 bp of the start codon. The 50bp oligomers were ordered as single stranded nucleotides but in complementary pairs. The full length *MEA*-ICR was amplified from wild-type genomic DNA in a standard PCR reaction (36 cycles of 94°C for 15 sec, 52°C for 20 sec, and 72°C for 30 sec; followed by 72°C for 5min; we used homemade Taq DNA Polymerase and PCR buffer and a final concentration of 2mM $MgCl_2$, 0.2mM dNTPs and 0.4mM Primer) using priMR1 and priMR2 (see Appendix A2) and purified by Phenol/Chloroform precipitation. For each binding assay 100 fmol of labeled probes are recommended. We 5' radiolabeled 2 pmol of probe (30ng for single-stranded 50 bp mer; 300 ng for double-stranded 250 bp mer) in 20 µl total volume by using T4 polynucleotide kinase (PNK, 20 units/reaction, New England Biolabs), [γ-^{32}P] ATP (2pmol, Hartmann Analytic), 1x T4 PNK buffer for 30min at 37°C (Maniatis et al., 1982). Then complementary single-stranded 50bp-mer oligonucleotides were mixed and all labeling reactions were incubated for 20 min at 65°C to stop the reaction. To allow the complementary oligonucleotides to anneal, the reactions were incubated at room temperature (RT) until cooling down.

Binding assay and PAGE

The radiolabeled probes (100fmol) were mixed with 2 µg bovine serum albumine (BSA), crude nuclear protein extract (20 to 80 µg, in an increasing manner) in 1x binding buffer (see Table 1-2-1, Binding Buffer B seemed to work best). To test binding specificity, a 100x excess of unlabelled probe was added to one additional sample containing the highest amount of protein sequestering a specifically bound protein. Furthermore, all reactions were done with or without 2 µg Poly(deoxyinosinic-deoxycytidylic) acid (Poly(dI-dC), Sigma Aldrich), a synthetic oligonucleotide sequestering unspecific DNA binding proteins. All binding reactions were incubated for

Table 1-2-1. Binding buffers tested and used in this study.

	stock	buffer A 5x[1]	buffer B 5x[2]	buffer C 5x[3]	buffer D 5x[4]	buffer E 5x[5]
Glycerol	100%	50%	50%	50%	50%	50%
Tris-HCl	1M	100mM		100mM	250mM	
HEPES	0.5M		125mM			50mM
NaCl	5M	100mM				
KCl	2.5M		250mM	250mM	500mM	300mM
$MgCl_2$	100mM	5mM		5mM		20mM
EDTA	0.5M		2.5mM	2.5mM	5mM	0.5mM
DTT	50mM	5mM	5mM	10mM	5mM	1.25mM
H_20						

1: Maniatis; 2: Urao et al. 1993; 3: own buffer; 4: Huang et al. 2004; 5: Sturm et al. 1987

30 min at RT, then loaded on a native polyacrylamide (PA, 6%, 0.5x TBE) gel and run for ~3-4h at 8-10V/cm in 0.5x TBE-buffer. Finally the PA gel was dried at 75°C for 1-2h and exposed over night using an X-ray film (Fuji Film) and developed the next day.

FORWARD GENETIC SCREEN

Plant material and growth conditions

The minimal *MEA*-ICR reporter gene (*250pMEA::GUS*) used for mutagenesis is published and described in detail (Wöhrmann et al., 2012). The line is in a gamma-ray irradiated C24 background of a wild-type segregant originating from a *dme-4/DME* population (Guitton et al., 2004), and we renamed this genetic background Fukushima (FUK, see below). The parental line used for backcrosses was the unmutagenized homozygous *250pMEA::GUS* line (HW 166.8 = MR 70). If required, the line was genotyped for the presence of the reporter construct in a standard PCR reaction using priMR9 and priMR10 to amplify the 730bp wild-type band and priMR9 and priMR11 to amplify the ~450bp mutant band (usually in a 3-primer-in-1-reaction, since the assay is robust; all primers used in this chapter are listed in the Appendix A2). Landsberg *erecta* (Ler) was used to cross the mutants out to create a mapping population. To screen for paternal repressor candidates, we used male-sterile *delayed-dehiscence2-2* (*dde2-2*, Col-0) mutant plants as mothers (von Malek et al., 2002). In the beginning we used an independent gene trap line (GT 27; *dde2-3*) in the Ler background that was generated by exploiting the *Ac/Ds* gene trap transposon tagging system (Sundaresan et al., 1995) within the EXOTIC consortium of the Fifth Framework program of the EU. Yet, we found ectopic GUS staining at the micropylar end of the seed in this line, which was independent of the paternal genotype and, therefore, might be due to the GUS gene in the gene trap construct (see below). Subsequently we used the original *dde2-2* line in Col-0 (von Malek et al., 2002) or emasculated FUK plants (wild-type segregants of *dme-4/DME*) to confirm transgene derepression in the paternal repressor candidates. For crosses, plants were emasculated and pollinated 2 days later to synchronize seed development and increase crossing efficiency. All plants were grown in a greenhouse chamber with 16h light at ~20°C and 8h darkness at ~18°C with an average of 60% humidity.

EMS mutagenesis

The differences between the mutagenesis of batch #1 and of batch #2 are listed in Table 1-2-2. In general, a batch of seeds (35'000-50'000 = 0.7-1.0 g) was hydrated in the dark for 4 days at 4°C, then re-dried for 24h at 22°C. The seeds were then immersed in a 0.2% EMS solution (20mM) for ~8h. After thorough washing,

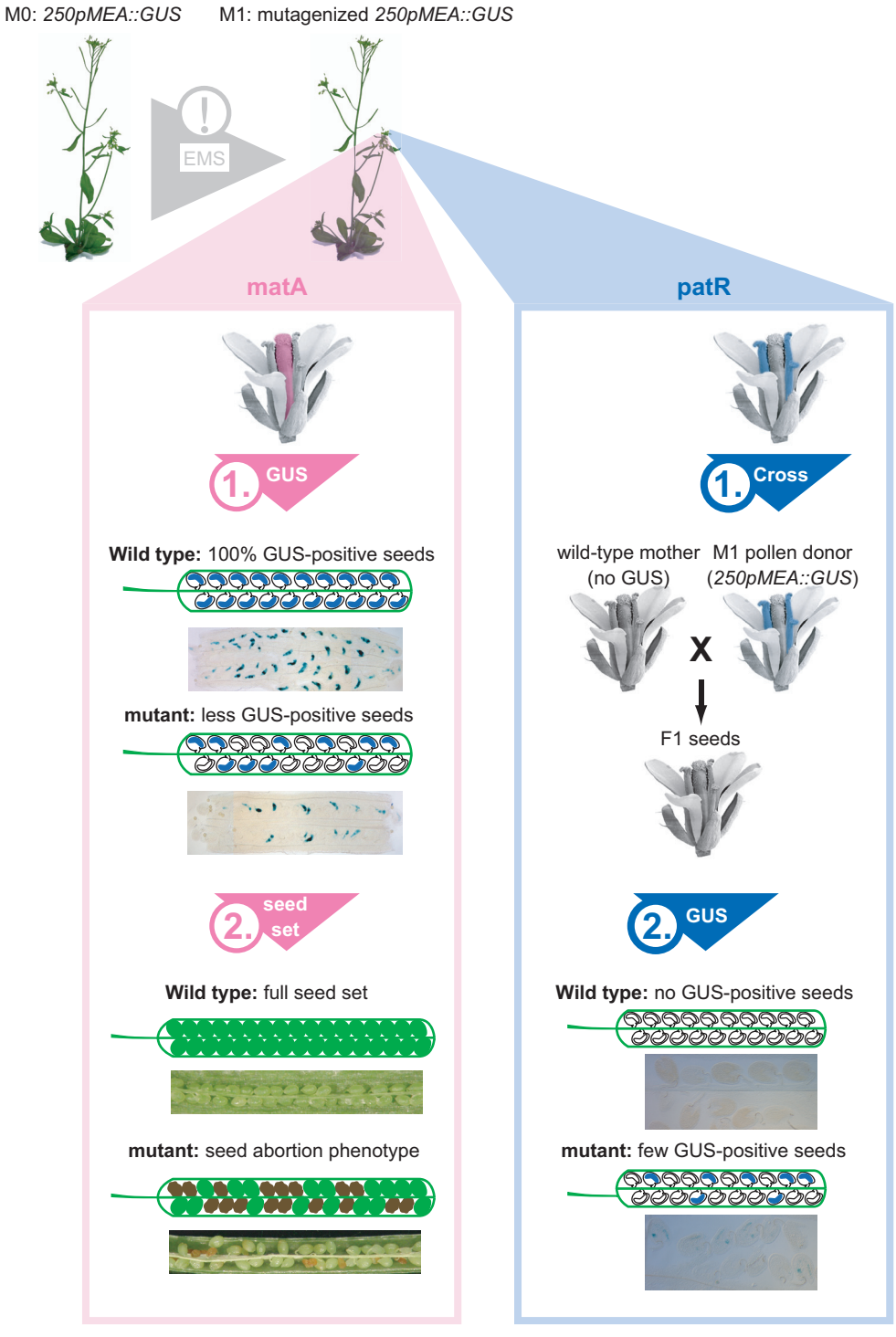


Figure 1-2-2. Screening scheme. Plants homozygous for the *250pMEA::GUS* reporter gene were mutagenized with 0.2% EMS. To screen for maternal activator mutants (matA, left, pink panel) we harvested and stained young M1 siliques and screened for a reduced number of GUS positive seeds. Then, we analyzed the seed set of all primary candidates showing aberrant GUS staining and focused on those that additionally displayed a seed abortion phenotype. To screen for paternal repressor mutants (patR, right, blue panel) we crossed M1 pollen to wild-type mothers, harvested and GUS-stained the resulting siliques and F1 seeds 2 DAP, and searched for ectopic expression of the usually silent paternal reporter gene.

the seeds were directly put on soil using the "salt-n-pepper" principle: A fraction of the mutagenized seeds (amount depending on the number of trays that should be planted) was resuspended in 50 ml water and filled in a 50 ml Falcon tube with holes in the lid and dispersed. We prepared 20-30 trays, germinated the plants and removed all but two clearly distinguishable plants per pot using forceps. At the rosette stage we put some trays (important: dry soil!) in a dark 4°C room to allow phasing of the screen, such that not all plants flower simultaneously but successively. We counted white sectors on M1 individuals and found that 1% of the plants showed a white sector, indicating successful mutagenesis (Jürgens et al., 1991).

Screening strategy, GUS staining and fertility phenotyping

The chimeric primary stem was removed and a single lateral branch per M1 plant was screened. We figured that we rather screen more M1 individuals than more sectors per M1 individual. The application of the screen in the M1 generation is possible, since regulation of imprinted *MEA* expression happens in the haploid gametophyte, where mutations have an immediate effect. To identify maternal activator candidates, we removed one silique at 1-2 days after pollination (DAP, batch #1) or at 0.5-1 DAP (batch #2, more consistent staining), cut it open with insulin-needles (longitudinal cuts or stem and stigma cuts), and immersed it in 4mM GUS staining solution (4 mM 5-bromo,4-chloro,3-indolyl-D-glucuronide (Biosynth-

Table 1-2-2. Comparison of the mutagenesis #1 and the mutagenesis #2.

		Screen #1	Screen #2
Mutagenesis	hydrating seeds	4d at 4°C on Whatman filters	4d at 4°C on fine mesh
	redrying	24h blotted on dry filter	24h blotted on dry filter
	[EMS]	0.20%	0.20%
	incubation time	8h	8.5h
	n(seeds)	0.7g = 35'000	1g=50'000
Population	n(plants)	896	1680
	plant line	MR70 = *250pMEA::GUS* (in C24)	MR70 = *250pMEA::GUS* (in C24)
	phasing	yes	yes
Screening	maternal activator	yes	yes
	paternal repressor	yes	no
	Time point	1-2dap (inconsistent)	0.5-1dap (more consistent)
	preparation of sq.	longitudinal cut + stigma cut	stigma + stem cut
	GUS staining time	48h	48h
	[GUS]	4mM	4mM
	Fixation, clearing	70% Ethanol	75% Ethanol
	dissection	yes (in Chloralhydrate)	no
	GUS phenotyping	microscope	binocular
	seed set analysis	no	yes

AG), 10 mM EDTA, 0.1% Triton X-100, 2.0 mM potassium ferrocyanide, 2.0 mM potassium ferricyanide, 50 mM phosphate buffer pH 7.2). After vacuum-infiltration for 5-15 min, we stained the samples for 48h at 37°C in 96-well, flat-bottom plastic plates. To stop the staining reaction, we removed the GUS staining solution and added 75% Ethanol to the siliques overnight to fix and clear the tissue. We then screened the siliques for reduced number of GUS-positive seeds, either after dissection and under the microscope (batch #1) or directly in the 96-well, flat bottom plate using the binoculars (batch #2). In batch #2, we screened all plants showing fewer GUS-positive seeds for a fertility phenotype by opening siliques, and assessed whether we find seeds arrested at the ovule stage (before or just at fertilization), at seed stages after fertilization, or full seed set and thus full fertility.

To identify paternal repressors (only batch #1) we crossed pollen of one sector of an M1 individual to male sterile mothers (*dde2-3;* GT27). F1 seeds were harvested 2 DAP and GUS stained as above. Seeds were dissected in clearing solution (8:2:1 chloralhydrate:glycerol:water) and analyzed using bright-field microscopy for ectopic, paternal *MEA*-like GUS expression in the fertilization products (i.e. endosperm and embryo). The screening scheme is taken together in Figure 1-2-2.

Histological analysis

To analyze the stage of embryo abortion in the mutant candidates we dissected seeds of heterozygous mutant individuals, separated aborted and plump seeds on a slide, and cleared them in modified Hoyer's solution (70% w/v chloralhydrate, 4% w/v glycerol, 5% w/v gum arabic). We then examined the stage and the patterning phenotype of mutant and wild-type seeds by differential interference contrast (DIC) microscopy (Leica DMR).

RESULTS

No Specific Band Shift Was Identified when Using EMSA

First of all, we tested different binding buffers, published or own buffers (Table 1-2-1), and found that binding buffer B (BB B) produced most band shifts with two differently assayed probes (225-175; 250bp *MEA-ICR*; Figure 1-2-3A). In a second step, we assayed all 50bp-mer oligonucleotides in a proper EMSA set-up with increasing protein concentration and an extra assay with 100x excess of cold, unlabeled competitor probe. Interestingly, we found band shifts for 256-200, 225-175 and 200-150, but no band shift was observed when using the 175-125 probe (Figure 1-2-3B). Yet, only the band shifts of 225-175 and the band shifts of 200-150 seemed to be due to a protein specifically binding to the probe, since the cold competitor was able to sequester most of the bound protein (Figure 1-2-3B), suggesting that potential regulatory proteins bind to the DNA sequence between -225bp and – 150bp of the start codon. This is in accordance with the work of Wöhrmann and colleagues, who found that the 50bp between -200bp and -150bp upstream of the start codon is required but not sufficient for *MEA*

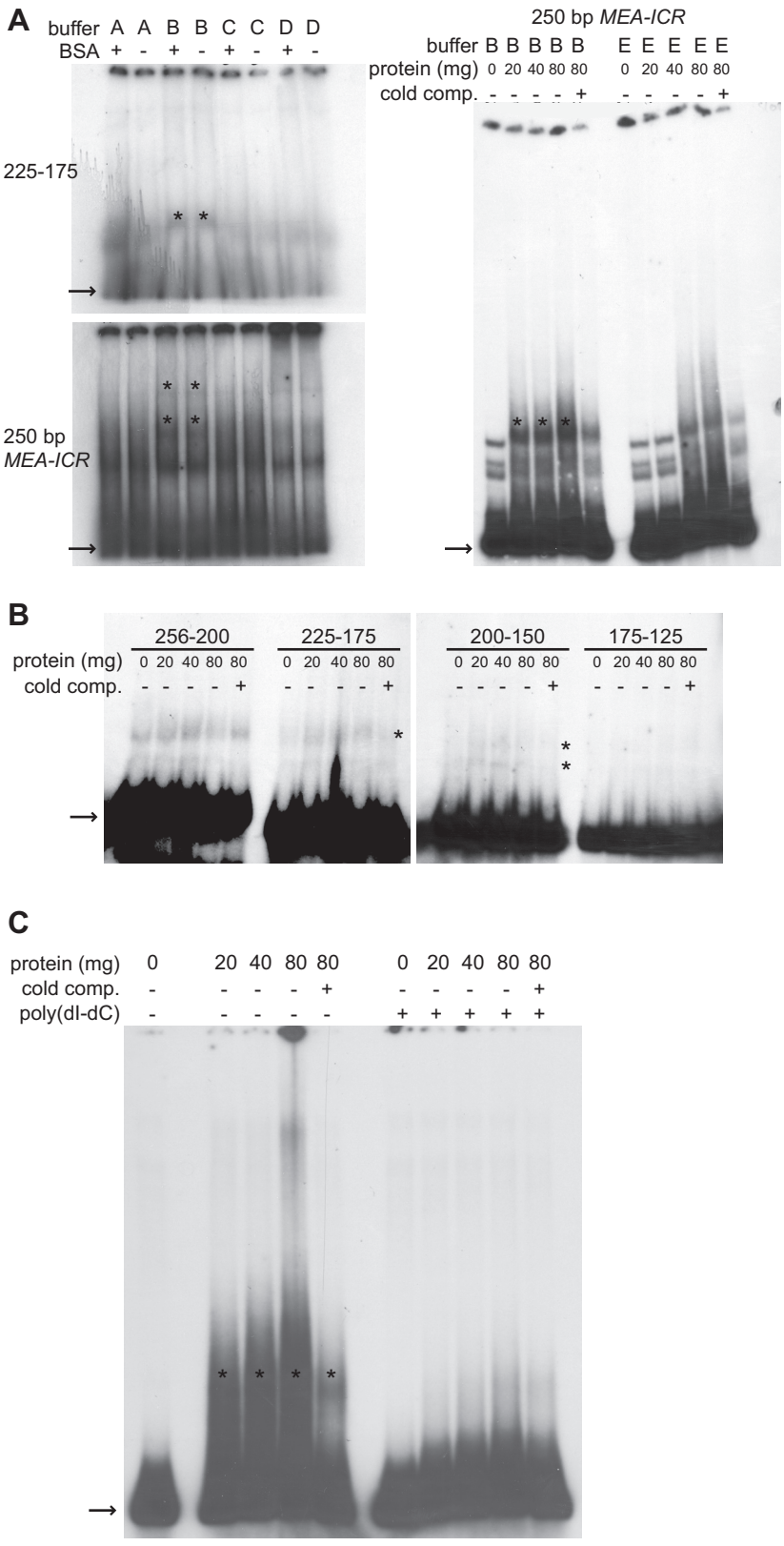


Figure 1-2-3. Band shift assay of radio-labeled fragments of the *MEA*-ICR with crude nuclear extracts from broccoli. (A) To test different binding buffers (see Table 1-2-1), EMSA was performed using either the full-length 250bp *MEA-ICR* probe (lower left panel, right panel) or the 225-175 probe. In the left panels the amount of added protein extract was constant and only buffer and addition of BSA varied as indicated on top of the panels. In the right panel added amount of protein was increased along the gel and cold competitor was added as indicated, as well as the binding buffer used. (B) EMSA of the four different oligonucleotides covering the regions of the *MEA-ICR* that are shown to be required for proper *MEA* expression. The oligonucleotide assayed, the amount of protein and the addition of cold competitor probe is indicated on top of each panel. Binding buffer B was used for all assays. (C) EMSA of the full-length 250bp *MEA-ICR* probe with or without the synthetic polymer poly(dI-dC). The amount of protein, the addition of the cold competitor probe, and the addition of the synthetic polymer poly(dI-dC) is indicated on top of each panel. Asterisks (*) indicate band shifts; arrows indicate the position of free, unbound probe.

expression (Wöhrmann et al., 2012).

However, when working with crude protein extracts it is absolutely indispensable to add the synthetic polymer poly(dI-dC), an alternating copolymer used as a DNA substrate, to the binding assays to prevent or reduce unspecific interactions of proteins with the labeled probe. By doing so, we found that all band shifts disappeared (Figure 1-2-3C), indicating that it was rather unspecific DNA binding proteins that stuck to the radiolabeled probes. We, therefore, were not able to consistently produce specific band shifts and abandoned this approach.

Identification of Two Maternal Activator Candidates in the Maternal Activator Screen

We screened a total of 2496 plants in two rounds after two independent EMS mutageneses. In the first batch (batch #1) we identified 18 candidates out of 889 screened M1 plants (~2%, Table 1-2-3 and Appendix A4.1). 14 candidates showed wild-type GUS staining (all seeds stained) in the M2 generation and were discarded, whereas two candidates (1-10V, 6-3V) did not show the expected phenotype after backcrossing, suggesting that the mutation is not transmittable via the pollen. Of the two remaining candidates, one showed an unexpectedly strong hybrid vigor growth habitus phenotype after backcrossing and was not followed up (3-19H), whereas the other one showed a clear reduction in GUS staining frequency (1-17H = matA1, Figure 1-2-4A) in the M2 generation. Any mutant candidate not activating the endogenous *MEA* gene should display reduced fertility since heterozygous mutant *mea* plants show a seed abortion frequency of 50%, and mutant seeds arrest at the heart stage (Grossniklaus et al., 1998). Yet, when assessing the seed set of matA1, we found that the seeds abort at the ovule stage just after fertilization or before fertilization (infertile ovules, Figure 1-2-4B). The endogenous *MEA* gene and the *MEA::GUS* reporter are expressed in the mature female gametophyte and early seed only (Grossniklaus et al., 1998, Wöhrmann et al., 2012). A mutant candidate that shows infertile ovules likely aborts female gametophyte development before *MEA::GUS* is expressed and is, therefore, likely a false positive. To test this hypothesis we cleared unfertilized ovules 3 days after emasculation and found that mutant ovules abort at the megaspore mother cell or the functional megaspore stage much before *MEA::GUS* expression (Figure 1-2-4C). In conclusion, we did not find promising maternal activator candidates in the first batch. We clearly underestimated the frequency of female gametophyte development arrest after EMS mutagenesis and, consequently, the occurrence of many false positives in this loss-of-function screen.

Table 1-2-3. Summary of identified maternal activator and paternal repressor candidates. For a detailed list, see Appendix A4.

		Screen #1	Screen #2
maternal activator	total plants screened	889	1603
	candidates (total)	18	240
	candidates (i.O.)	?	97
	candidates (i.O. and s.a.)	?	85
	candidates (s.a.)	?	40
	candidates (no s.a.)	?	17
	confirmed candidates (BC2)	0	**2**
paternal repressor	total plants screened	783	n.d.
	candidates	17	n.d.
	confirmed candidates (BC2)	**2**	n.d.

To reduce the retrieval of false positive candidates, we adjusted the screening strategy for the second batch (batch #2): We first screened, as before, for a reduced GUS staining frequency in young seeds and then assessed the seed set of all previously identified mutant candidates in a second step (see Figure 1-2-2). We screened 1603 plants and identified a total of 240 primary mutant candidates displaying an aberrant GUS staining phenotype (15%,

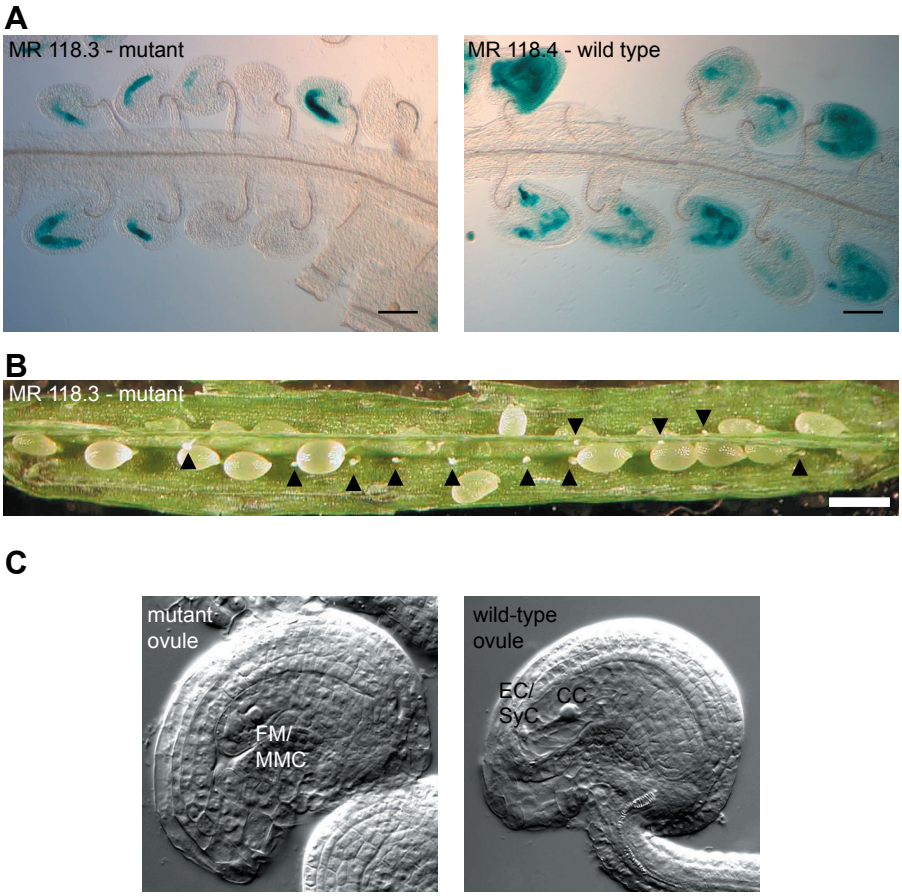


Figure 1-2-4. Phenotype of Maternal Activator 1 (matA1). (A) GUS staining phenotype of a mutant (left panel, MR118.3) and a wild-type individual (right panel, MR118.4) in the M2 generation. GUS staining intensity and frequency are clearly reduced in the mutant individuals. (B) matA1 mutant individuals display a reduced seed set and infertile ovules (black arrowheads). (C) Clearing of mutant and wild-type ovules shows that the female gametophyte aborts during its development, namely at the megaspore mother cell (MMC) or functional megaspore (FM) stage. Wild-type ovules clearly display a fused central cell nucleus (CC) and an egg cell (EC) or synergid cell (SyC) nucleus. Black scale bar = 50 µm; white scale bar = 1 mm.

Table 1-2-3 and Appendix A4.3-A4.7). By assessing the seed set we could exclude 75% of all primary mutant candidates that showed either infertile ovules only, or a combination of infertile ovules and aborted seeds after fertilization. Of the remaining 59 candidates, 40 showed aborted seeds as we would expect for a mutant candidate that does not activate *MEA*, and 19 showed a wild-type seed set, including 2 candidates that showed ectopic *MEA::GUS* expression in the sporophyte. We decided to rescreen all 40 candidates showing seed abortion, and all 17 candidates showing aberrant GUS staining frequencies but full fertility. We hypothesized that any candidate showing aborted seeds might be involved in the establishment of *MEA* activation whereas any fully fertile candidate might be rather maintaining the active state of *MEA* after fertilization. We rescreened all candidates either directly in the F1 of the first backcross (BC1) or, if the backcross was not available, in the M2 generation. We discarded all candidates, which now showed a normal, wild-type GUS staining pattern (n=33), which showed infertile ovules instead of aborted seeds, or full seed set (n=24), and which displayed an unlinked GUS-staining and fertility phenotype (n=45, see Appendix A4.5 and A4.6). Rescreening resulted in three promising candidates displaying seed abortion and reduced GUS staining: maternal activator 15 (matA15), matA18 and matA25 (Appendix A4.5 and A4.6). Unfortunately, we found in the BC2 population of matA15 (n=144) that the GUS phenotype and the fertility phenotype are not linked and, therefore, discarded this candidate. Interestingly, we could not recover a single candidate that showed a reduced GUS staining frequency in the seed but full fertility, suggesting that there is no maintaining activator among the population screened.

In the end, we identified two transmittable mutant candidates, which showed a reduction in reporter gene activation accompanied by late seed abortion: matA18 and matA25.

MatA18 and MatA25 Display Reduced Reporter Expression and Seed Abortion

The two identified maternal activator candidates, matA18 and matA25, showed a similar phenotype in terms of *MEA::GUS* reporter expression and fertility. We saw a clear reduction of GUS positive seeds in both mutants (Figure 1-2-5A and 1-2-5C). In addition to the absolute reduction of reporter expression, we found

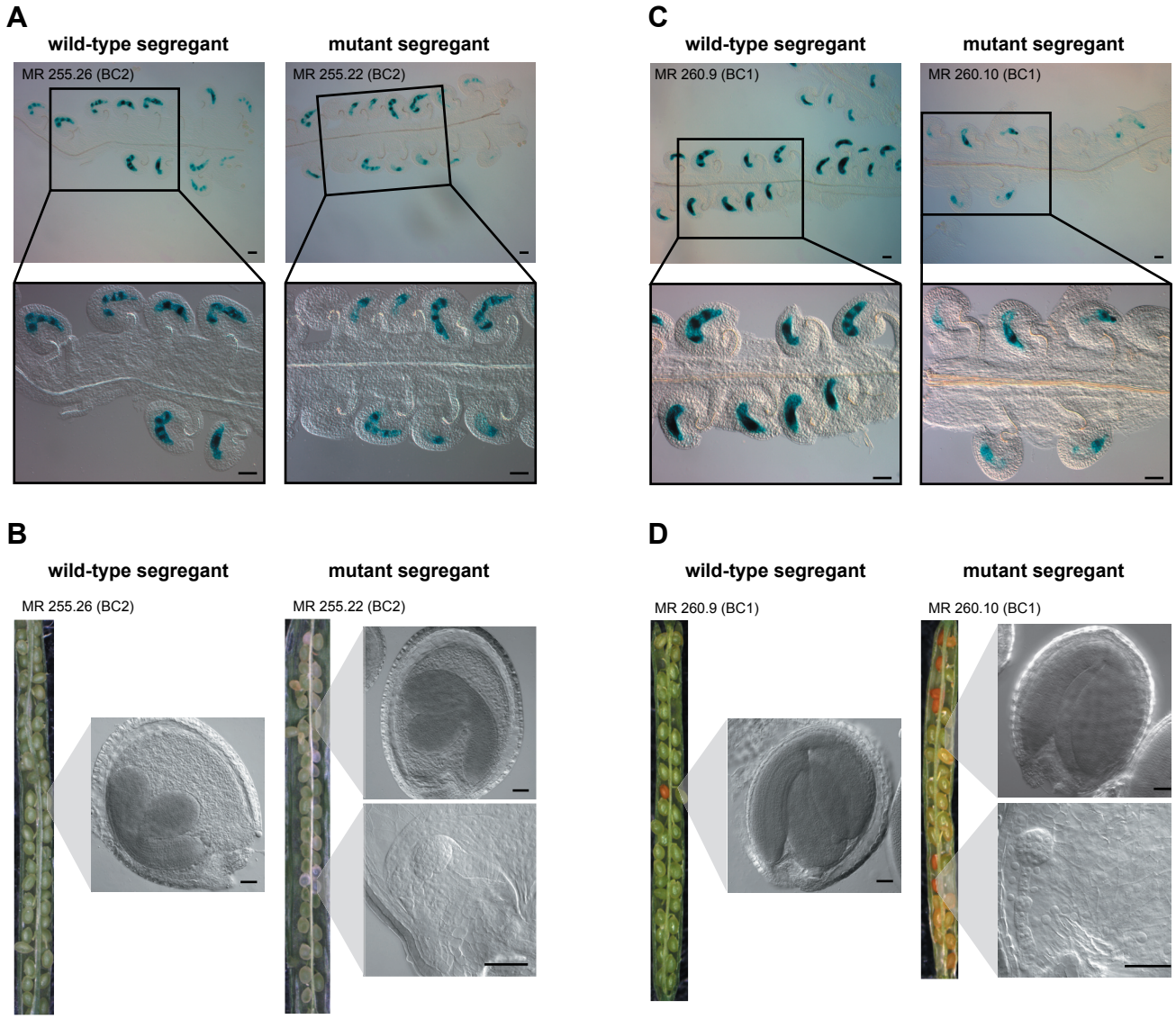


Figure 1-2-5. Phenotype of the maternal activator candidates, matA18 and matA25. (A) and (C) GUS phenotype of matA18 (A) and of matA25 (C). Whereas in the wild-type individuals (MR255.26; MR260.9) all seeds express the GUS reporter, the mutant individuals of matA18 (MR255.22) and matA25 (MR 260.10) display a reduction of reporter expression, both in frequency and intensity. (B) and (D) Fertility phenotype of matA18 (B) and of matA25 (D). Wild-type individuals (MR255.26; MR260.9) are fully fertile and show a full seed set with nicely developing embryos. Mutant individuals (MR255.22; MR 260.10) show plump and aborted seeds in both maternal activator mutants. The mutant seeds abort around the globular embryo stage and the embryo shows abnormal cell division planes and ectopic cell divisions. Scale bar = 50 µm.

seeds that displayed a differential staining pattern, where a fraction of the seeds clearly and consistently showed weaker reporter expression than in a wild-type silique (Figure 1-2-5A and 1-2-5C). Although this is rather difficult to quantify, we tried to do so in three mutant individuals compared to three wild-type individuals in the BC5 of matA18: We found that ~85% of all seeds in mutant siliques still expressed the GUS reporter (n=314), but that 84 of the 267 GUS positive seeds clearly expressed the GUS reporter at much weaker levels than wild-type seeds in the same silique (31%). Thus, a complete loss of GUS expression in 15% of the seeds and an additional 31% of seeds expressing the reporter at much weaker levels in heterozygous mutants, suggested either a polygenic mutant, a developmental timing effect, or a dosage effect. Yet, the mutation was readily transmitted to the next generation with male transmission frequencies of 51% (n=338) and 44% (n=322) for matA18 and matA25, respectively (Appendix A4.8 and A4.9). The female transmission efficiency was only assessed for matA18 and was 73% (n=96, Appendix A4.8 and A4.9). The high transmission frequency of both mutations suggests that the mutant is likely not of polygenic origin or that the disrupted genes must be closely linked. In addition, the GUS staining

pattern of the other maternal activator candidate, matA25, was very similar although we did not quantify it. The probability to disrupt twice the same combination of genes by chance is close to impossible, especially in a small screening population of only 2500 M1 plants. Therefore, the differential staining phenotype likely arises due to a dosage effect of the mutation or is impaired in developmental timing.

As mentioned before, we expected that non-activation of *MEA* should result in *mea*-like fertility defects. In fact, we found that both mutants abort seed development around globular stage (Figure 1-2-5B and 1-2-5D). Interestingly, embryo patterning and cell division was impaired in both mutants. Microscopic analysis of cleared seeds revealed wrong and chaotic embryonic cell division planes, including ectopic cell divisions in the suspensor and tumor-like outgrowth in the embryo proper (Figure 1-2-5B and 1-2-5D). In addition, we quantified the seed abortion frequency and found that 22.9% ± 2.4% of the seeds in matA18 abort and that 22.4% ± 6.4% of the seeds in mat25 abort. A seed abortion frequency of approximately 25% suggests a recessive nature of the observed embryo lethality, which is rather unexpected. Yet, the mutant seeds abort much earlier than *mea*-mutant seeds (globular vs. late heart stage) suggesting that not only *MEA* is affected in our mutant candidates and that the seed abortion might be due to an epistatic effect. In contrast, the observed seed abortion frequency of 25% could also be due to the dosage-dependent activation of *MEA::GUS* and probably of the endogenous *MEA* leading to abortion of the fraction of the seeds, which are below a certain threshold of *MEA* expression or, in which the activation of *MEA* is outside of the required developmental time frame. Lastly, we observed a small fraction of infertile ovules in mutant siliques, especially in matA25 (4.9%), which is either due to the causative mutation and, again, dependent on a dosage effect of impaired *MEA* activation or due to additional EMS mutations or the genetic background. In fact, we did not observe infertile ovules in all mutant individuals (~40%, n=67) hinting rather at an unlinked background mutation.

In summary, we identified two maternal activator candidates, matA18 and matA25. Both candidates show a reduction in *MEA::GUS* reporter expression in the seed, both quantitative and qualitative, and a quarter of the seeds abort in heterozygous mutants, suggesting an embryo-lethal mutant effect on seed development or a dosage-dependent effect. We aimed at mapping the mutations using our recently developed SRM approach (see Chapter 1.3; Lindner et al., 2012).

Mapping of MatA18 and MatA25 by SRM Reveals a Highly Variable Genetic Background

SRM is based on the distinct segregation ratio of the causative (and linked) SNP from that of unlinked SNPs: After backcrossing twice to the non-mutagenized parent, any unlinked, EMS-induced SNP is diluted and, thus, segregates 1:3 in a pool of individuals. By selecting only mutant individuals in the BC2, the causative SNP is enriched such that it segregates 1:1 in a pool of mutant BC2 individuals (see Chapter 1.3; Lindner et al., 2012). Hence, we phenotyped a BC2 population of matA18 and matA25 (n=222 and n=215, respectively), extracted DNA of all mutant individuals and pooled the genomic DNA of 49 mutant matA18 individuals and 69 mutant matA25 individuals after DNA extraction in an equimolar fashion. At the same time we isolated DNA from the unmutagenized parental line for resequencing (*250pMEA::GUS*, FUK, *dme4/DME* wild-type segregant) to distinguish EMS-induced SNPs and polymorphisms due to the genetic background.

For resequencing of the parental and unmutagenized *250pMEA::GUS* line, we prepared three independent samples: We sequenced twice a pool of three individuals (250MEA-1 and 250MEA-2) and we sequenced once a single individual (250MEA-3). Not surprisingly, we called many SNPs when mapping the reads to the Col-0 reference genome. Yet, depending on the parental sample we found different numbers of SNPs: In the two samples where we pooled 3 individuals, we found 349'350 SNPs and 151'283 SNPs in 250MEA-1 and 250MEA-2, respectively (Table 1-2-4). In addition, many of the SNPs were heterozygous, which was very unexpected in a theoretically clean and inbred ecotype. Furthermore, in the sample, which consisted of a single individual

Table 1-2-4. Overview table of called SNPs of the parental and the mutant lines mapped against different reference genomes. 250MEA_1 and 250MEA_2 are two independently pooled (n=3), unmutagenized parental line samples (*250pMEA::GUS*). 250MEA_3 corresponds to a single individual of the unmutagenized parental line (*250pMEA::GUS*). matA18 and matA25 are the two mutant lines. Col-0 corresponds to the fully annotated and published reference genome (TAIR10, www.arabidopsis.org), C24 (1001) is the Col-0 genome with the incorporated SNPs of C24 as published on www.1001genomes.org, C24_FGCZ_v1 is the Col-0 genome with the incorporated SNPs that were called from the resequencing run of the parental line, and C24_FGCZ_v2 is C24_FGCZ_v1 plus all identified INDELs incorporated. Shown in the table are the called SNPs in total, the homozygous and the heterozygous SNPs and the called INDELs of -3bp to +10bp. n.d. not done. Numbers in italic are approximate.

			250MEA_1 n=	250MEA_2 n=	250MEA_3 n=	matA18 n=	matA25 n=
Col-0	SNPs	total	349'337	151'270	99'554	432'050	407'289
		homo	322'443	141'109	93'280	396'727	368'779
		hetero	26'894	10'161	6'274	35'323	38'510
	INDELs	total	43'033	23'533	17'492	54'817	55'381
		homo	31'591	19'633	15'275	34'228	32'656
		hetero	11'442	3'900	2'217	20'589	22'725
C24 (1001)	SNPs	total	192'465	77'024	49'880	n.d.	n.d.
		homo	167'120	70'475	46'168	n.d.	n.d.
		hetero	25345	6'549	3'712	n.d.	n.d.
	INDELs	total	47'404	29'510	22'365	n.d.	n.d.
		homo	35'410	24'832	19'718	n.d.	n.d.
		hetero	11'994	4'678	2'647	n.d.	n.d.
C24_FGCZv1	SNPs	total	57'819	22'263	12'402	139'389	n.d.
		homo	31'587	11'976	6'169	106'767	n.d.
		hetero	26'232	10'287	6'233	32'622	n.d.
	INDELs	total	43'033	24'486	17'998	58'256	n.d.
		homo	31'591	20'424	15'716	36'518	n.d.
		hetero	11'442	4'062	2'282	21'738	n.d.
C24_FGCZv2	SNPs	total	*50'000*	n.d.	n.d.	139'329	n.d.
		homo	*20'000*	n.d.	n.d.	108'617	n.d.
		hetero	*30'000*	n.d.	n.d.	30'712	n.d.
	INDELs	total	*13'000*	n.d.	n.d.	31'429	n.d.
		homo	*4'000*	n.d.	n.d.	12'952	n.d.
		hetero	*9'000*	n.d.	n.d.	18'477	n.d.

only (250MEA-3), but was sequenced together with the pooled 250MEA-2, we called only two thirds of the SNPs (n= 99'567) when compared to the 250MEA-2 sample (n=151'283, Table 1-2-4). This suggests, that the individuals differ between each other, which again was not expected in an inbred parental line. Nevertheless, we then mapped all 250MEA samples against a Col-0 reference genome that was corrected *in silico* for C24 polymorphisms by incorporation of the mapped and annotated C24 SNPs of the 1001 genome project (C24_1001; www.1001genomes.org). Here, we found less SNPs, but still 192'478, 77'037, and 49'893 in 250MEA-1, 250MEA-2 and 250MEA-3, respectively, suggesting that the accession used in this study dramatically differs from the published C24 accession (Table 1-2-4). Lastly, we incorporated all newly identified SNPs of the resequencing runs into the Col-0 reference genome to create the *250pMEA::GUS* reference genome *in silico* and to map the reads of the mutant candidates against the new and adapted reference genome (C24_FGCZ_v1). To control the quality of our newly created *250pMEA::GUS* reference genome, we again mapped the 250MEA-1 sample against it: Regardless, we still found many SNPs (n= 57'819), especially many heterozygous SNPs (n= 26'232) when we used the corrected *250pMEA::GUS* reference genome (C24_FGCZ_v1) and whether we incorporated small INDELS (-3bp until +10bp; C24_FGCZ_v2) or not (Table 1-2-4).

Nevertheless, we sequenced and mapped the mutant candidate matA18 to the corrected *250pMEA::GUS* reference genome (C24_FGCZ_v1) and identified more than 106'000 homozygous SNPs and over 32'000 heterozygous SNPs (see Table 1-2-4). In fact, we expected a small number of EMS-induced, heterozygous SNPs in the mutagenized line plus many homozygous, ecotype-specific SNPs that should be masked when mapping against a corrected reference genome on the basis of the resequenced and unmutagenized parental line. At the same time of sequencing matA18, we sequenced and mapped a different EMS-induced mutant candidate gene, *turan* (*tun*), in Col-0 using SRM (see Chapter 1.3; Lindner et al., 2012). We found 521 homozygous SNPs (likely lab-strain variation polymorphisms) and 1816 heterozygous SNPs, which are likely induced by the applied EMS mutagenesis (see Chapter 1.3; Lindner et al., 2012). The conservatively estimated mutation frequency per locus per diploid cell and per dose is $3.7 \pm 0.5 * 10^{-6}$ per 10mM EMS (Koornneef et al., 1982). Since we use ~20-30mM EMS per mutagenesis, we would expect ~1000 to 1500 EMS induced SNPs in a mutant line, a number that is consistent with what we observed when mapping *tun* by SRM. Yet, for matA18, we observed a very high number of polymorphisms: After mapping the matA18 reads against the corrected C24_FGCZ_v1, we found 20 times more heterozygous SNPs than in the *tun* mutant. In addition, there were inconsistencies when comparing the three independent samples of the resequenced parental line.

Therefore, we investigated the origin of our line and its genetic background. We found that the original line designated as C24, which we used to transform the *250pMEA::GUS* reporter, was a wild-type segregant of a *dme-4/DME* population (Guitton et al., 2004). The *dme-4* mutation was isolated from a forward genetic screen after gamma-ray irradiation, a procedure known to induce many large and small INDELs and point mutations (Wu et al., 2005). In addition, all mutant lines were backcrossed only four times (Guitton et al., 2004), which is not sufficient to remove all unlinked SNPs, explaining some of the segregating heterozygous SNPs in the sequenced lines. Given that gamma-ray irradiation causes much fewer lesion than EMS however, we do not understand where the larger number of SNPs stems from. Lastly, wild-type segregants of epigenetic mutants may have an altered epigenetic landscape that might be maintained even after the mutant allele was segregated away. Thus, working with wild-type segregants of an epigenetic mutant like *dme-4*, when studying a process like genomic imprinting that is epigenetically regulated in angiosperms and mammals, bears some intrinsic difficulties and should be omitted. Due the highly distinct genetic background of the C24 strain used in this study compared to the published C24, we decided to rename this ecotype: The time of the discovery that we are working in a heterogenous genetic background, which was previously gamma-ray irradiated, coincided with the meltdown of the nuclear reactor in Fukushima, Japan, in spring 2010 and, therefore, prompted us to call the background henceforth FUKUSHIMA (FUK).

As a last and rather desperate approach, we tried to simultaneously sequence and map both maternal activator lines, matA18 and matA25, to the fully annotated, high quality reference genome Col-0 (TAIR 10). This parallel approach would allow to subtract common variants and could result in a list of unique, EMS-induced SNPs per mutant candidate if the genetic background is homogenous (Zuryn et al., 2010). We identified 351'609 SNPs that are common to both mutant candidates, but still identified 80'441 unique SNPs (61'214 homozygous SNPs and 19'227 heterozygous SNPs) in matA18 and 55'681 unique SNPs (40'093 homozygous SNPs and 15'588 heterozygous SNPs) in matA25. The identification of almost 20'000 unique heterozygous SNPs in matA18 and more than 15'000 unique heterozygous SNPs in matA25 are an order of magnitude more than we expected from the literature (Koorneeff et al., 1982) and from what we found when applying SRM on the mutants of Lindner and colleagues (see Chapter 1.3; Lindner et al., 2012; Lindner, Grossniklaus, unpublished). Furthermore, the presence of approximately 60'000 and 40'000 unique homozygous SNPs further indicated that the parental line was far from being inbred and being genetically homogenous. In addition the mean read coverage of the heterozygous SNPs is only 31x and 19x for matA18 and matA25, respectively, making SRM, which required a

minimal coverage of 50x, very difficult. The low coverage is likely caused by the high degree of polymorphisms between reads and reference genome, reducing mapping efficiency and, thus, coverage. All together, there are still many polymorphisms segregating in the genetic background of the FUK accession, making the mapping of our mutant candidates by SRM impossible. We do not know whether the variable genetic background of our line is due to the gamma-ray irradiation, the insufficient rounds of backcrosses after irradiation, the fact that we deal with a wild-type segregant of an epigenetic mutant, or a combination of all.

Finally, we decided to backcross the two maternal activator lines at least five times to Col-0 to create near-isogenic lines (NILs). Five backcrosses should statistically lead to 96.875% pure Col-0 sequence and 3.125% of residual FUK sequence at and around the causative SNP. We just obtained the F1 seeds of the 5th backcross and could confirm the phenotype in all generations suggesting that the mutant phenotype is due to an EMS-induced mutation and not due to a combinatorial, polygenic effect of the FUK background. We aim to resequence the maternal activator NILs to quickly pinpoint the region of the mutation.

Taken together, we were not able to map the maternal activator candidates, matA18 and matA25, by SRM. The genetic background of the mutagenized *250pMEA::GUS* line was affected by a prior gamma-ray irradiation and subsequently not sufficiently backcrossed, creating segregating SNPs and strong differences on the genome level between different individuals of the same population. In addition, the line was transformed into a wild-type segregant of a *dme-4* mutant population, a mutation known to affect the epigenetic landscape and, thus, potentially genomic imprinting in plants even after being segregated away. Naturally, we stopped the screen at that point and only continued to create the NILs and, eventually, would like to try mapping these promising candidates again.

Identification of Two Paternal Repressor Candidates, PatR2 and PatR9

To screen for paternal repressor candidates we crossed M1 pollen to male-sterile mother plants supposedly not carrying any GUS reporter genes and analyzed the F1 seeds for derepressed reporter expression (Figure 1-2-2). Usually, the *250pMEA::GUS* line is paternally silent throughout seed development (Wöhrmann et al., 2012). If we disrupt any paternal repressor acting in the male gametophyte before fertilization we expect to observe ectopic, paternal *250pMEA::GUS* expression in the endosperm or the embryo.

We performed the paternal repressor screen by screening batch #1 only and identified 17 primary candidates. We followed up on the three candidates that derepress the paternally silent reporter gene the most: paternal repressor 2 (patR2 = 2-20H), patR8 (8-13V), and patR9 (9-8H, Appendix A4.2). We crossed the M1 generation to *dde2-3* (Ler) male-sterile mother plants and found an additional ectopic staining at the micropylar end, which was independent on the paternal genotype (Figure 1-2-6C). Thus, we decided to assess the derepression phenotype in the following generations on different mother plants, such as emasculated wild-type FUK segregants of a *dme-4/DME* population or *dde2-2* (Col-0) mother plants that do not show any additional GUS signals. In the BC1 generation we found derepressed GUS expression in 30-50% of the resulting seeds in all three paternal repressor candidates (Appendix A4.2). In contrast to patR2 and patR9, where the paternal derepression of the usually silent *250pMEA::GUS* line was easy to score and clearly visible (Figure 1-2-6A and 1-2-6B), this was not case for patR8: The derepressed GUS reporter was expressed at too low levels for consistent scoring and, therefore, we did not follow up the patR8 candidate any further. Although the paternal GUS signal of patR2 and patR9 was clear and easily visible, the frequency of mutant individuals per population, the proportion of stained seeds per silique, and the intensity of derepression were varying a lot between the generations (Appendix A4.2). To test whether different lines used as mothers have an effect on imprinted expression of the *250pMEA::GUS* reporter gene, we crossed non-mutagenized *250pMEA::GUS* reporter line pollen to different mother plants, including Col-0, *dde2-2* (Col-0), FUK (wild-type segregants of *dme-4/DME* a.k.a. C24), and C24 (NASC, a newly

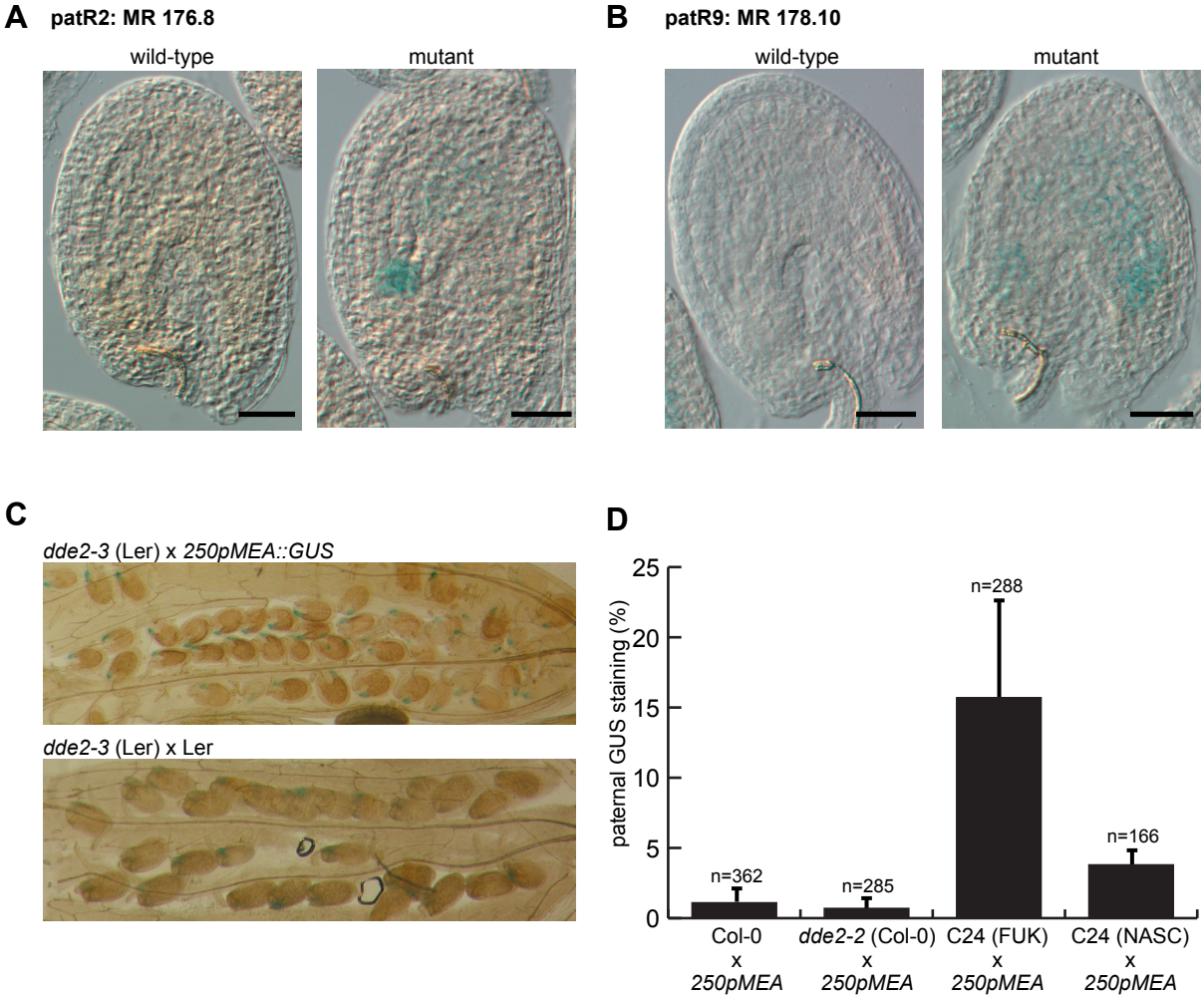


Figure 1-2-6. Phenotype of the paternal repressor candidates, patR2 and patR9. (A) and (B) If the seed inherited a wild-type genome from the pollen (left panel) no reporter expression can be detected, but when a mutant paternal genome was inherited, a *MEA*-like GUS signal is clearly visible in both paternal repressor candidates, patR2 (A) and patR9 (B). (C) The micropylar GUS signal of *dde2-3* is visible independently of the paternal genotype. If inheriting pollen devoid of any GUS reporter gene (Ler), the micropyle is stained. The staining likely derives from the gene trap construct that disrupts the *DDE2* gene in Ler (GT27). (D) Paternal *250pMEA::GUS* expression in different maternal backgrounds. Whereas the reporter is clearly imprinted and silent in Col-0 and *dde2-2* (Col-0) mothers, the reporter is expressed in 15% of all seeds in C24 (FUK) mothers but in less than 5% of all seeds in C24 (NASC stock center ecotype). The genotype of the cross is indicated above or below each panel; sample size is indicated where appropriate (n=); scale bar = 50 µm.

ordered C24 line from the NASC stock center). As expected, we hardly found any GUS signal when the reporter was crossed to Col-0 and *dde2-2* (Col-0), suggesting that the reporter remains imprinted in those genetic contexts (Figure 1-2-6D). Yet, when we crossed the non-mutagenized reporter line to FUK wild-type segregants we found expression of the reporter line (~16%), which we did not find when using stock-center-derived C24 mother plants (Figure 1-2-6D). This suggests that either the genetic (due to heavy gamma-ray irradiation) or the epigenetic landscape (due to the former presence of the *dme-4* mutation) of the FUK wild-type segregants is altered and influences the expression of the *250pMEA::GUS* reporter. Likely, the paternal derepression phenotype of patR2 and patR9 relies solely on the genetic or epigenetic background of the original line. We might not have pulled out original, monogenic candidates, but rather combinatorial, polygenic traits in the FUK background. Therefore, we decided not to continue with the paternal repressor candidates. Especially in the light of the sequencing results (see above) with thousands of SNPs that sometimes even segregate between siblings, we did not trust anymore our paternal repressor candidates and decided to abandon the paternal repressor screen. If, at all, we either propose to start the screen all over again, or to cross out the paternal repressors to Col-0 plants to create NILs as we did for the maternal activator candidates.

DISCUSSION

To identify novel regulators of imprinted *MEA* expression we attempted (i) a biochemical screen (EMSA) and (ii) a forward genetic screen: The biochemical screen was quickly abandoned, since there are too many uncontrollable aspects to this assay. We worked with crude nuclear extracts, surely having a high abundance of DNA-binding proteins that might mask a specific interaction. Yet, fractionation or other enrichment techniques of the crude extract might have helped to enhance specific interactions. Furthermore, we used crude nuclear extracts of *B. oleracea* var. *italica* and not of *Arabidopsis*, which itself might be suboptimal. First, we cannot predict whether a potential regulatory protein of *B. oleracea* var. *italica* would indeed bind to the *Arabidopsis cis* elements. Second, we do not know whether *MEA* is imprinted in *B. oleracea* var. *italica*, and third, if it is imprinted, it might not be controlled by the same regulators in *trans* and/or the same elements in *cis*. In addition, EMSA is a very sensitive assay, which produces rather variable results depending on the binding buffers used, the amount of the synthetic polymer poly(dI-dC) added, and the way the proteins are isolated, fractionated or enriched. Lastly, we could not estimate whether isolation of a specific interaction was possible and sufficient, in terms of amount and specificity for mass spectrometry analysis. We thus decided to focus on the genetic approach and abandon the biochemical screen.

The forward genetic screen for novel regulators of imprinted *MEA* expression yielded in two maternal activator candidates, matA18 and matA25, both of which display reduced reporter expression in the seed and abort seeds at the globular stage. We tried to map the causative mutation using SRM, a method relying on whole-genome resequencing and the distinct segregation ratio of linked and unlinked SNPs in a pool of BC2 mutant individuals. However, we identified way too many SNPs, homozygous and heterozygous, to map the loci using SRM. In addition, the resequencing of the parental line, which was done to create a decent reference genome of our parental background, revealed inconsistencies between different samples of the same line, many INDELS, and, strangely, a huge population of segregating, heterozygous SNPs that should not be present in an inbred and unmutagenized parental line. We discovered that the original C24 line used for transformation with the *250pMEA::GUS* reporter construct was a wild-type segregant of a mutant *dme-4/DME* population. In addition to potential hazardous effects on studying imprinting regulation in a wild-type segregant of an epigenetic mutant, the *dme-4* mutation was identified in a screen following mutagenesis by gamma-ray irradiation. Lastly, the mutants were crossed back only four times, an insufficient number to remove all background polymorphisms.

Nevertheless, even if the history of this line explains some of the identified variability and polymorphism, a single mutagenesis event cannot account for more the 100'000 SNPs, in particular after gamma-ray irradiation , which causes fewer lesions than EMS. A further possibility is that the mutant candidates are impaired in DNA repair mechanisms and display a MUTATOR-like behavior (Lisch, 2002). Yet, if this would be the case, we would (i) find much less SNPs in the parental line and (ii) observe accumulating mutant phenotypes over time and generations after inbreeding. However, in this background, we were not able to map the mutation. In addition, we could not know whether the identified maternal activator candidates are really based on the disruption of a single gene by EMS mutagenesis or whether they are rather dependent on the FUK background. Crossing out matA18 and matA25 to Col-0 at least showed that the mutation is transmittable at relatively high rates, indicating that indeed a single locus is responsible for the phenotype. In future, we will try to map and clone the genes in the F1 of the 5th backcross to Col-0, likely by whole genome resequencing to quickly pinpoint the polymorphic region and then identify the causative SNP by various filtering strategies (see below).

The two identified paternal repressor candidates, patR2 and patR9, consistently derepress the paternal *250pMEA::GUS* reporter gene when crossed as fathers. Yet, the frequency, the intensity and the transmission of the mutant allele varied tremendously between different generations of the same paternal repressor candidate.

Furthermore, the imprinted status of the *250pMEA::GUS* reporter line was partly lost when crossed to FUK wild-type mother plants, suggesting that some processes do not function properly in the FUK accession. Regarding the unexplainable accumulation of polymorphisms in the FUK accession, together with the inconsistencies concerning reporter gene derepression between generations and the loss of the imprinted status of the *250pMEA::GUS* reporter on FUK wild-type mother plants, we attributed the phenotype of the paternal repressor candidates rather to its FUK background than to a monogenic, EMS-induced mutation.

Nonetheless, we identified many additional maternal activator candidates that bear some potential to be interesting and that we did not follow up in the scope of this thesis. For instance in the group of maternal activator candidates showing both, seed abortion and infertile ovules, we probably wrongly declared some candidates as false positives. The infertile ovules could arise from secondary mutations not linked to the mutant effect on *250pMEA::GUS* reporter expression and could be segregated away. In addition, there might be false negative candidates that show infertile ovules only. We did not clear the ovules and assess the stage of abortion and, therefore, we can speculate that we would find candidates with mature, aborted ovules, or very young seeds just after fertilization. Lastly, there are four additional candidates that showed interesting phenotypes, but were not followed up: First, we identified two mutant candidates that lost tissue specificity of the *250pMEA::GUS* reporter and showed strong staining in the seed coat (10-13H and 15-4V, Appendix A4.7). In these two candidates, either a sporophytic repressor was mutated, which is rather unlikely, since in the M1 generation all mutations are heterozygous and, therefore, recessive traits would not be noticed, or we mutated a *cis* element in the reporter construct. Second, there were two additional maternal activator candidates confirmed in the M2 generation. Both showed reduced GUS staining frequency in the seed, but candidate 24-7V has a full seed set indicating that a factor involved in maintaining expression of *MEA* or the reporter gene itself could be affected, whereas candidate 24-13V shows aborted seeds and infertile ovules like matA25 (Appendix A4.5 and A4.6). Both candidates were confirmed only in the M2 generation, but the BC1 generation is available.

Although speculating about the possible function of the maternal activator genes is of no avail yet, we would like to stress the presence of five DNA BINDING WITH ONE FINGER (DOF)-binding sites within the *MEA*-ICR (Chapter 1.1; Wöhrmann et al., 2012). DOFs are DNA-binding proteins containing a C2C2-type Zinc-finger-like motif and there are 37 putative DOF transcription factors in the *Arabidopsis* genome that regulate diverse biological processes like defense or seed germination (Yanagisawa, 2002 and references therein). Of course DNA-binding transcription factors like DOFs could activate *MEA* by binding to the *MEA*-ICR and are, therefore, eligible candidates. Nevertheless, we are not even close to map the maternal activator candidates, making the presence of the DOF binding sites our only hint towards the functional nature of the maternal activator but remains purely speculative.

In conclusion, we tried to identify novel regulators of imprinted MEA expression biochemically and genetically. EMSA bears too many intrinsic problems, such as unspecific binding or binding buffer composition, and we quickly abandoned this approach. Yet, if protein extracts of slightly higher quality (eg. fractionated) and if the right tissue and stage are used, and smaller and more specific DNA elements probed, this method might proof useful in future. The forward genetic screen produced promising mutant candidates. Unfortunately, we did not realize the problems with the genetic background of the *250pMEA::GUS* line early enough, and found a tremendous excess of polymorphisms. This made mapping by SRM impossible. Currently, we prepare NIL populations of two maternal activator candidates in order to attempt mapping of the causative genes again.

FUTURE DIRECTIONS

With the availability of matA18 and matA25 as NILs in a Col-0 background, the mapping of the two mutant alleles should be tried again. We propose to resequence the NIL genome at relatively low coverage to make it affordable, map it to the Col-0 reference genome and identify the region with many SNPs. In the region around the causative SNP, there should be thousands of heterozygous SNPs that derive from the remaining FUK background due to a low chance of recombination events close to the causative SNP. However, with the SNP list at hand, it is possible to sort the SNPs for exonic, non-synonymous SNPs in first place, and in a second step, sort the SNPs for the usual G/C to A/T transition induced by EMS (Sega, 1984). This would reduce the list of candidate SNPs tremendously (see Chapter 1.3; Lindner et al., 2012). In addition, all causative mutations identified by Lindner and colleagues produced a premature stop codon (Lindner, Grossniklaus, unpublished). Therefore, filtering for stop codons might be a further benefit. Finally, the best candidates could be found according to the predicted function and when taking available gene expression data into account.

In addition to a pure genetic approach, a molecular Yeast-1-Hybrid screen would be a further possibility to find proteins that interact with the *MEA*-ICR. Obviously, a bait construct that is not self-activating and that has a few repeats of the *MEA*-ICR upstream of a selection gene is indispensible and needs to be cloned. In addition, a decent screening library needs to be produced. We propose to use either whole inflorescences, young flower buds, or even the *apetela1* (*ap1*) *cauliflower* (*cal*) double mutants carrying an inducible *AP1* rescue construct (Wellmer et al., 2006). *ap1 cal* double mutant plants arrest (temporarily) flower formation and overproliferate inflorescence-like meristems (Wellmer et al., 2006). Upon induction of *AP1,* these plants simultaneously produce massive amounts of floral buds, the stage where *MEA* activation takes place (Wellmer et al., 2006; Grossniklaus et al., 1998). Therefore, using this system would facilitate the collection and generation of the required tissue at the required stage. In fact, this system would not only be suitable to produce a Y1H screening library, but also to extract proteins for any further EMSA attempts.

Lastly, the identification of the key sequence within the *MEA*-ICR is of great interest and would facilitate the construction of a bait construct for Y1H assays and potentially simplify further biochemical attempts to find interacting partners by applying EMSA. A linker scanning mutagenesis technique (McKnight and Kingsbury, 1982), where specific sequences within the *MEA*-ICR are replaced by different sequences of the same length, was initiated in our laboratory. This approach will hopefully pinpoint the (short) imprinting control element (ICE) within the *MEA*-ICR.

References

Folta, K.M. and Kaufman, L.S. (2006). Isolation of *Arabidopsis* nuclei and measurement of gene transcription rates using nuclear run-on assays. Nat Protoc **1**: 3094–3100.

Garner, M.M. and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic acids Res **9**: 3047–3060.

Grossniklaus, U., Vielle-Calzada, J.P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal control of embryogenesis by *MEDEA*, a *polycomb* group gene in *Arabidopsis*. Science **280**: 446–450.

Guitton, A.E., Page, D.R., Chambrier, P., Lionnet, C., Faure, J.E., Grossniklaus, U., and Berger, F. (2004). Identification of new members of Fertilisation Independent Seed *Polycomb* Group pathway involved in the control of seed development in *Arabidopsis thaliana*. Development **131**: 2971–2981.

Jürgens, G., Mayer, U., Ruiz, R.T., Berleth, T., and Mlséra, S. (1991). Genetic analysis of pattern formation in the *Arabidopsis* embryo. Development Supplement **1**: 27–38.

Koornneef, M., Dellaert, L.W.M., and van der Veen, J.H. (1982). EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) *Heynh*. Mutat Res **93**: 109–123.

Lindner, H., Raissig, M.T., Sailer, C., Shimosato-Asano, H., Bruggmann, R., and Grossniklaus, U. (2012). SNP-Ratio Mapping (SRM): Identifying Lethal Alleles and Mutations in Complex Genetic Backgrounds by Next-Generation Sequencing. Genetics **191**: 1381–1386.

Lisch, D. (2002). Mutator transposons. Trends Plant Sci **7**: 498–504.

von Malek, B., van der Graaff, E., Schneitz, K., and Keller, B. (2002). The *Arabidopsis* male-sterile mutant *dde2-2* is defective in the *ALLENE OXIDE SYNTHASE* gene encoding one of the key enzymes of the jasmonic acid biosynthesis pathway. Planta **216**: 187–192.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

McKnight, S.L. and Kingsbury, R. (1982). Transcriptional control signals of a eukaryotic protein-coding gene. Science **217**: 316–24.

Sega, G.A. (1984). A review of the genetic effects of ethyl methanesulfonate. Mutat Res **134**: 113–142.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D., Dean, C., Ma, H., and Martienssen, R. (1995). Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev **9**: 1797–810.

Wellmer, F., Alves-Ferreira, M., Dubois, A., Riechmann, J.L., and Meyerowitz, E.M. (2006). Genome-wide analysis of gene expression during early *Arabidopsis* flower development. PLoS Genet **2**: e117.

Wu, J.-L., Wu, C., Lei, C., Baraoidan, M., Bordeos, A., Madamba, M.R.S., Ramos-Pamplona, M., Mauleon, R., Portugal, A., Ulat, V.J., Bruskiewich, R., Wang, G., Leach, J., Khush, G., and Leung, H. (2005). Chemical- and Irradiation-induced Mutants of Indica Rice IR64 for Forward and Reverse Genetics. Plant Mol Biol **59**: 85–97.

Wöhrmann, H.J., Gagliardini, V., Raissig, M.T., Wehrle, W., Arand, J., Schmidt, A., Tierling, S., Page, D.R., Schöb, H., Walter, J., and Grossniklaus, U. (2012). Identification of a DNA methylation-independent imprinting control region at the *Arabidopsis MEDEA* locus. Genes Dev **26**: 1837–1850.

Yanagisawa, S. (2002). The Dof family of plant transcription factors. Trends Plant Sci **7**: 555–560.

Zuryn, S., Le Gras, S., Jamet, K., and Jarriault, S. (2010). A strategy for direct mapping and identification of mutations by whole-genome sequencing. Genetics **186**: 427–430.

PART 1
CHAPTER 3

SNP-RATIO MAPPING (SRM): IDENTIFYING LETHAL ALLELES AND MUTATIONS IN COMPLEX GENETIC BACKGROUNDS BY NEXT-GENERATION SEQUENCING

NOTE

All of CHAPTER 1.3 is published as **Lindner, H.*, Raissig, M.T.*, Sailer, C., Shimosato-Asano, H., Bruggmann, R., and Grossniklaus, U.** (2012). SNP-Ratio Mapping (SRM): Identifying Lethal Alleles and Mutations in Complex Genetic Backgrounds by Next-Generation Sequencing. Genetics **191**: 1381–1386.

**these authors contributed equally to this work*

MTR contributed to this work as follows:

HL and MTR published this work in co-first authorship and developed the method together with UG. The method was applied to map mutant candidates isolated from two different forward genetic screens aimed at identifying (i) regulators of imprinted *MEA* expression (see Chapter 1.2), and (ii) additional members of the *FERONIA (FER)* signaling pathway required for pollen tube reception. The available next-generation sequencing based mapping procedures, like the SHOREmap approach, (Schneeberger et al., (2009) *Nature Methods* **6**, 550 - 551), require that a mutation can be isolated in the homozygous state and rely on a large population, two conditions that cannot be met with our mutations. Thus, we developed SRM, and successfully mapped 3 mutant candidates in the Col-0 background isolated by HL (Chapter 1.3; Lindner, Grossniklaus, unpublished) but were not able to map the maternal activator candidates, likely due to the FUK background (Chapter 1.2). We published the successful mapping of *TURAN* (*TUN*), a putative UDP-glycosyltransferase superfamily protein involved in pollen tube reception. Therefore, HL is mentioned first in the author list. Supplemental information to this publication can be found in Appendix A5.

SNP-Ratio Mapping (SRM): Identifying Lethal Alleles and Mutations in Complex Genetic Backgrounds by Next-Generation Sequencing

Heike Lindner,*[,1] Michael T. Raissig,*[,1] Christian Sailer,* Hiroko Shimosato-Asano,*[,2] Rémy Bruggmann,[†,‡] and Ueli Grossniklaus*[,3]

*Institute of Plant Biology and Zürich-Basel Plant Science Center, University of Zürich, CH-8008 Zürich, Switzerland, [†]Functional Genomics Center Zürich of the Eidgenössische Technische Hochschule and University of Zürich, CH-8057 Zürich, Switzerland, and [‡]Interfacultary Bioinformatics Support, Department of Biology, University of Bern, CH-3012 Bern, Switzerland

ABSTRACT We present a generally applicable method allowing rapid identification of causal alleles in mutagenized genomes by next-generation sequencing. Currently used approaches rely on recovering homozygotes or extensive backcrossing. In contrast, SNP-ratio mapping allows rapid cloning of lethal and/or poorly transmitted mutations and second-site modifiers, which are often in complex genetic/transgenic backgrounds.

FORWARD genetic screens are powerful in uncovering novel gene functions in genetic model organisms. While some mutant screens can be quick to perform, the identification of the causative mutation by map-based cloning is extremely labor-intensive. Large F_2 mapping populations of >1000 mutant individuals are required (Lukowitz *et al.* 2000; Jander *et al.* 2002) to fine-map a chromosomal region harboring a causative mutation. This number of mutant individuals can be difficult to obtain, especially when working with phenotypic traits that (i) are difficult to score, (ii) are weakly transmitted, or (iii) are in organisms that are hard to propagate. The recent development of next-generation sequencing (NGS) platforms has made sequencing of whole genomes quick and affordable. One application of NGS is to replace map-based cloning by the sequencing of mutagenized genomes to quickly identify causative mutations, a method successfully applied in many model organisms (Sarin *et al.* 2008; Smith *et al.* 2008; Srivatsan *et al.* 2008; Blumenstiel *et al.* 2009; Irvine *et al.* 2009; Schneeberger *et al.* 2009; Zuryn *et al.* 2010; Austin *et al.* 2011). However, current methods depend on identifying homozygous mutant individuals in an F_2 mapping population after outcrossing (Schneeberger *et al.* 2009; Austin *et al.* 2011) or require several rounds of backcrossing (Zuryn *et al.* 2010), a time-consuming requirement not easily met in organisms with long generation times.

Here, we describe a generally applicable method, SNP-ratio mapping (SRM), which allows the rapid identification of lethal and/or poorly transmitted mutations and second-site modifiers by NGS. It is based on the distinct segregation ratio of the causative (and linked) single-nucleotide polymorphism(s) (SNPs) from that of unlinked SNPs. SRM allows the mapping of lethal mutations after only two rounds of backcrossing via NGS. After backcrossing twice to the non-mutagenized parent, any unlinked SNP created by ethyl methanesulfonate (EMS) mutagenesis segregates 1:3 in a pool of individuals. By selecting only mutant individuals in the F_1 generation of the second backcross (BC2), the causative SNP is enriched and segregates 1:1 in a pool of mutant BC2 individuals (Figure 1). Thus, calculating the SNP/non-SNP segregation ratio allows the quick identification of the causative mutation. The method is applicable to any model organism and mutagen causing mostly point mutations or small indels. SRM is the method of choice when working with (i) lethal mutations, (ii) hard-to-score phenotypes, (iii) mutations with low transmission, and (iv) second-site modifiers in complex genetic/transgenic backgrounds. Here, we demonstrate the power of

doi: 10.1534/genetics.112.141341
Manuscript received February 9, 2012; accepted for publication May 21, 2012
Supporting information is available online at http://www.genetics.org/content/suppl/2012/05/26/genetics.112.141341.DC1.
[1]These authors contributed equally to this work.
[2]Present address: Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan.
[3]Corresponding author: Institute of Plant Biology and Zürich-Basel Plant Science Center University of Zürich, Zollikerstrasse 107, CH-8008 Zürich, Switzerland. E-mail: grossnik@botinst.uzh.ch

SRM by cloning a gametophyte lethal mutation in *Arabidopsis thaliana*, for which the recovery of homozygotes is not possible.

As proof of principle, we aimed to map the gene affected in a pollen-tube reception mutant obtained from a forward genetic screen using EMS-treated seeds of *A. thaliana* (Col-0 accession, Supporting Information, File S1). The *turan-1* (*tun-1*) mutant disrupts cell–cell communication between male and female gametophytes, which is indispensable for fertilization. In flowering plants, the gametes are produced by the haploid, multicellular gametophytes. The male gametophyte (pollen tube) delivers two sperm cells to the female gametophyte (embryo sac), harboring two female gametes. Fertilization of the egg and central cell forms the embryo and the endosperm, respectively. In heterozygous *tun-1* mutants, 12% (n = 1318 ovules) of the embryo sacs remain unfertilized, compared to only 1.5% (n = 1389 ovules) in the wild-type control. In *tun-1* mutants, the pollen tube fails to stop growing inside the female gametophyte and does not rupture to release the sperm cells, which leads to a pollen-tube overgrowth phenotype revealed by aniline-blue staining of callose in the pollen tube's cell wall (Figure 2 and File S1). Due to impaired fertilization and an additional effect of the mutant in the pollen, the transmission of the mutation is highly reduced, and homozygous individuals cannot be recovered. Thus, recently published methods for mutant allele identification by NGS (Schneeberger *et al.* 2009; Austin *et al.* 2011) are not applicable to mapping this gametophyte lethal mutation.

To identify the *TUN* gene by SRM, heterozygous mutants were crossed back twice to the wild-type Col-0 parent. By selecting only mutant individuals in the F_1 generation of the BC2, the causative SNP is enriched and segregates 1:1 in a pool of mutant BC2 individuals, whereas any unlinked SNP segregates 1:3 (Figure 1). We simulated a binomial distribution for a 1:1 and a 1:3 segregation to determine the optimal sample size and calculated that a 50-fold sequence coverage of the *Arabidopsis* genome was sufficient to distinguish a SNP segregating 1:1 from a SNP segregating 1:3 ($P < 0.05$, Table S1). Genomic DNA from 53 F_1 individuals of the BC2 generation that displayed the mutant phenotype was pooled for sequencing (File S1). A sequencing library was prepared (File S1) and sequenced on the SOLiD 4 platform, as this method provides an incomparable sequencing accuracy optimal for SNP detection. Reads were mapped to the *A. thaliana* genome assembly and SNPs were called and analyzed (File S1).

We identified 2337 SNPs, of which 521 were homozygous and 1816 were heterozygous with an average sequence coverage of 57 reads (Table S2 and Table S3). The homozygous SNPs were likely due to discrepancies between our lab strain of Col-0 and the published sequence. The homozygous SNPs were discarded, since all relevant SNPs should only be heterozygous (Figure 1). Before plotting the SNP/non-SNP ratios of the heterozygous SNPs, we filtered any SNPs that showed very low or high coverage. Low-coverage SNPs could exhibit a misleading ratio due to small sample size, while very high coverage (>2× average coverage) SNPs often mapped to repetitive and/or transposable element sequences, where mapping quality is usually poor (Figure S1). Thus, we filtered out the lowest (< 19×) and the highest (> 103×) 10% quantiles, leaving 80% of the original data set. The SNP/non-SNP ratio of the remaining 1468 heterozygous SNPs was calculated and plotted against their chromosomal position (Figure 3). Any unlinked SNP should have a SNP/non-SNP ratio of $\sim$ 0.25, whereas a causative SNP is expected to segregate 1:1, *i.e.*, producing a SNP/non-SNP ratio of 0.5. Furthermore, the SNPs surrounding the causative mutation should have segregation ratios > 0.25 since they have been coselected and thus cosegregate due to genetic linkage. Using this method, the causative SNP can be easily identified on the basis of the criteria that it must have a segregation ratio of $\sim$ 0.5, while the flanking, noncausative SNPs should cosegregate and display a ratio between 0.25 and 0.5, depending on the genetic/physical distance. The closer the flanking SNPs are, the higher this ratio will be. On the segregation ratio plot, this results in a rounded, rather flat curve, which can be visually identified without further statistical analyses (Figure 3A, red shading). Noncausative SNPs with a segregation ratio of 0.5 are likely to be surrounded by SNPs with low segregation ratios, leading to sharp drops in the SNP/non-SNP ratios of nearby SNPs (Figure 3). In our analysis of the *tun-1* mutant, the only rounded peak was present in the upper arm of chromosome I (Figure 3A, red shading).

Although the causative SNP could easily be identified in our experiment, we did not want to rely on a visual identification of the rounded peak. Thus, we developed a statistical test based on the expected recombination rate of neighboring SNPs as a function of the genetic distance between the SNPs. For each SNP following the 1:1 binomial distribution (n = 118, coverage $\geq$ 50), we calculated the expected pattern of cosegregation with the two neighboring SNPs on each side by using the expected recombination rate according to the mean genetic distance of 1 cM/357,042 bp (File S1 and File S2). Using a χ^2 goodness-of-fit test, 108 of 118 1:1 class SNPs did not lie in a linkage group (5 neighboring SNPs) that fit the expected pattern of cosegregation and therefore were discarded. Of the 10 remaining candidate SNPs, 8 reside in recombination-deficient centromeric regions. This is probably due to intrinsic problems in mapping reads to the highly repetitive centromeric sequences, leading to a high SNP density with unusual segregation ratios. Moreover, these eight linkage groups encompass 0.013 cM or less (Table S4), and the probability that 5 random SNPs lie in such close proximity is $P = 1.4 \times 10^{-8}$ (Poisson distribution, $\lambda = 0.071$, $k = 5$). Thus, any 5 SNPs that are in such close vicinity are likely of artificial nature due to mapping errors and should not be considered. In contrast, the linkage groups of the two remaining noncentromeric SNPs cover a genetic distance of 4.7 cM (ratio = 0.49) and 6.8 cM (ratio = 0.44), respectively. Both SNPs lie in the rounded peak that we visually

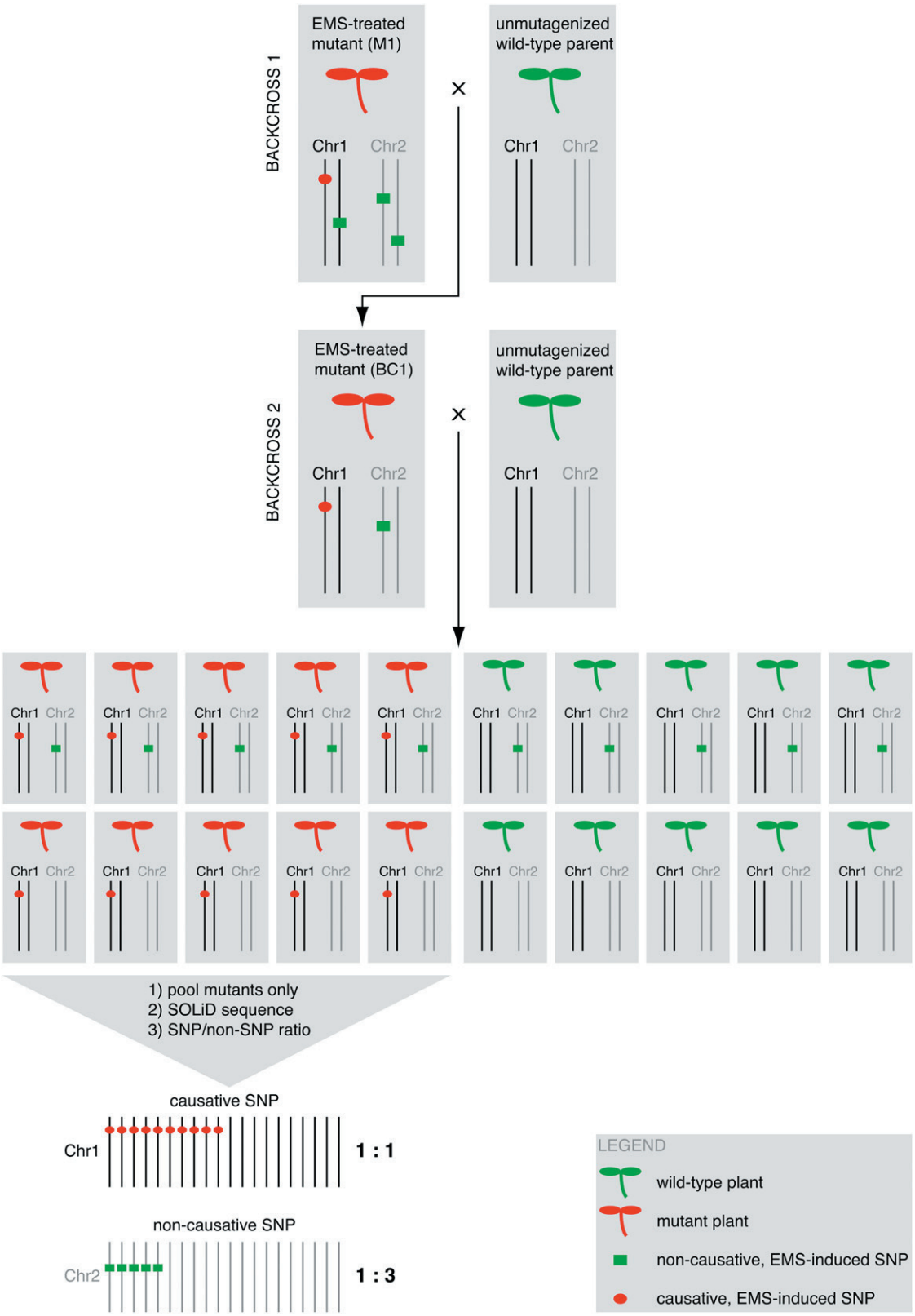


Figure 1 SRM scheme. An EMS-treated mutant (red plant) harboring several EMS-induced SNPs throughout the genome is backcrossed to an unmutagenized wild-type parent (green plant) of the same accession. The first backcross eliminates half of the SNPs. The F_1 generation is phenotyped, and a single mutant individual (red plant, BC1) is backcrossed. The F_1 of the second backcross (BC2) is phenotyped, and genomic DNA of 25–50 mutant individuals is extracted and pooled. The causative SNP (red circle) is present in every mutant individual in a heterozygous state and thus segregates 1:1 in a pool of mutant individuals. Unlinked SNPs (green square) are not selected and thus segregate 1:3. After SOLiD sequencing, SNP calling, and SNP/non-SNP ratio calculation, the two different segregation ratios can be distinguished.

Note **1383**

identified on the upper arm of chromosome I and are neighbors (Figure 3, red shading).

Of the two visually and statistically identified 1:1 class SNPs, the SNP with a ratio of 0.44 was intronic whereas the SNP with a segregation ratio of 0.49 (the closest to 0.5 in the whole data set) (Figure 3A, arrow) was a nonsynonymous GC-to-AT nucleotide change, which is characteristic of most EMS-induced SNPs (Sega 1984). This nucleotide change produces a stop codon in the sixth exon of gene *At1g16570*, a putative UDP-glycosyltransferase superfamily protein. To demonstrate that the causative SNP was identified, the *At1g16570* gene was amplified from each of the 53 DNA samples that had been pooled for sequencing (File S1). The PCR products were digested with nucleases cleaving single-base-pair mismatches in heteroduplex DNA (Till *et al.* 2004). Using this method, 52 samples were shown to have a SNP at the indicated position, while one sample was not cut (Figure S2). The progeny of this plant showed no phenotype, indicating that it was a sampling mistake due to wrong phenotyping in the BC2 generation. Finally, T-DNA insertion lines disrupting the identified gene *At1g16570* were tested for a pollen-tube reception phenotype. The line SAIL_400_A01 (*tun-2*), which has an insertion in the fourth exon of *At1g16570*, displays the same pollen-tube overgrowth phenotype as the EMS allele *tun-1* (Figure 2C), indicating that the correct gene has been identified by SRM.

In this example, no further analyses were required to identify the causative SNP. However, if genome coverage or mapping quality of the reads is lower than expected, the application of several filtering strategies could narrow down the list of potential candidate SNPs. First, selecting for exonic SNPs (nonsynonymous, synonymous) removes most of the detected SNPs (Table S2). Second, prioritizing characteristic EMS-induced SNPs (Sega 1984) should unambiguously identify the causative SNP in most cases. If not, then the rare cases where the mutation affects a regulatory region that is not exonic or an atypical EMS-induced nucleotide change have to be considered.

Interestingly, we also observed SNP ratios > 0.5. Since this should not be possible considering our genetic backcrossing strategy (Figure 1), we performed a detailed analysis of all SNPs on chromosome I with ratios > 0.5. All such SNPs display a low coverage (low-sample-size effect) or are covered by reads with low mapping quality (Figure S1). This indicates that such high ratios might be mapping artifacts in repetitive and/or transposable element regions. In addition, any SNPs found in and around the centromere display unusual segregation ratios (Figure 3, gray shading), probably representing an intrinsic problem in mapping sequence reads to highly repetitive centromeric regions.

On the whole, labor-intensive, map-based cloning has been replaced by cloning via NGS in recent years. Until now, this worked only (i) with homozygous viable mutants using the SHOREmap or similar strategies (Schneeberger *et al.* 2009; Austin *et al.* 2011) after outcrossing or (ii) for organisms with a short generation time and an easy-to-score phenotype, where multiple required backcrosses still save time (Zuryn *et al.* 2010). In contrast, SRM enables the mapping of zygotic or even gametophytic lethal mutations after only two rounds of backcrossing. SRM is generally applicable, but the identification of heterozygous mutant individuals may require progeny tests, *e.g.*, scoring for the presence of aborted seeds or defective embryos among the progeny. This might involve some adaptations of the crossing scheme shown in Figure 1, including a combination of inter se crosses and backcrosses if selfing is not possible. In outcrossing species, individual males could first be crossed to siblings to identify the heterozygotes in a progeny test, as well as to wild-type females to generate the backcrossed progeny used for SRM. In fact, SRM could also be used for the cloning of the causative genes on the basis of homozygous mutants in F_2 populations: the expected SNP ratios would be different (1.0 *vs.* 0.5), but the approach would still benefit from the small number of individuals required.

SRM is especially advantageous for (i) lethal mutations, (ii) organisms with a long generation time, (iii) hard-to-score phenotypes, and (iv) mutations with low transmission because only a small number of individuals are needed. Importantly, SRM is the method of choice for second-site modifier screens, in which a mutant with a certain phenotype is mutagenized a second time to identify novel mutant alleles that enhance or suppress this phenotype. Again, classical mapping or the SHOREmap strategy (Schneeberger *et al.* 2009; Austin *et al.* 2011), which rely on outcrossing and an F_2 mapping population, require the original mutation to be present in at least two genetic backgrounds. This is possible only when another allele is available in a different accession or by outcrossing the mutation five to six times to another accession. This procedure is time-consuming and has the disadvantage that, due to a lack of recombination

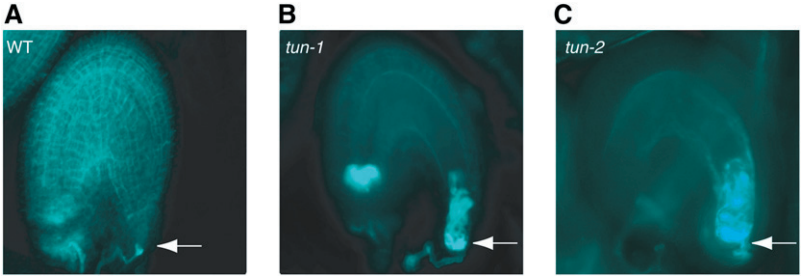


Figure 2 Aniline-blue staining of callose in pollen tubes 2 days after pollination. The arrow indicates the place of pollen-tube arrest. (A) Fertilized wild-type ovule. (B) Ovule harboring a *tun-1* embryo sac with defective pollen-tube reception. The pollen tube continues its growth and does not rupture to release the sperm cells. (C) Pollen-tube overgrowth phenotype in *tun-2*, an independent T-DNA line disrupting the *At1g16570* gene.

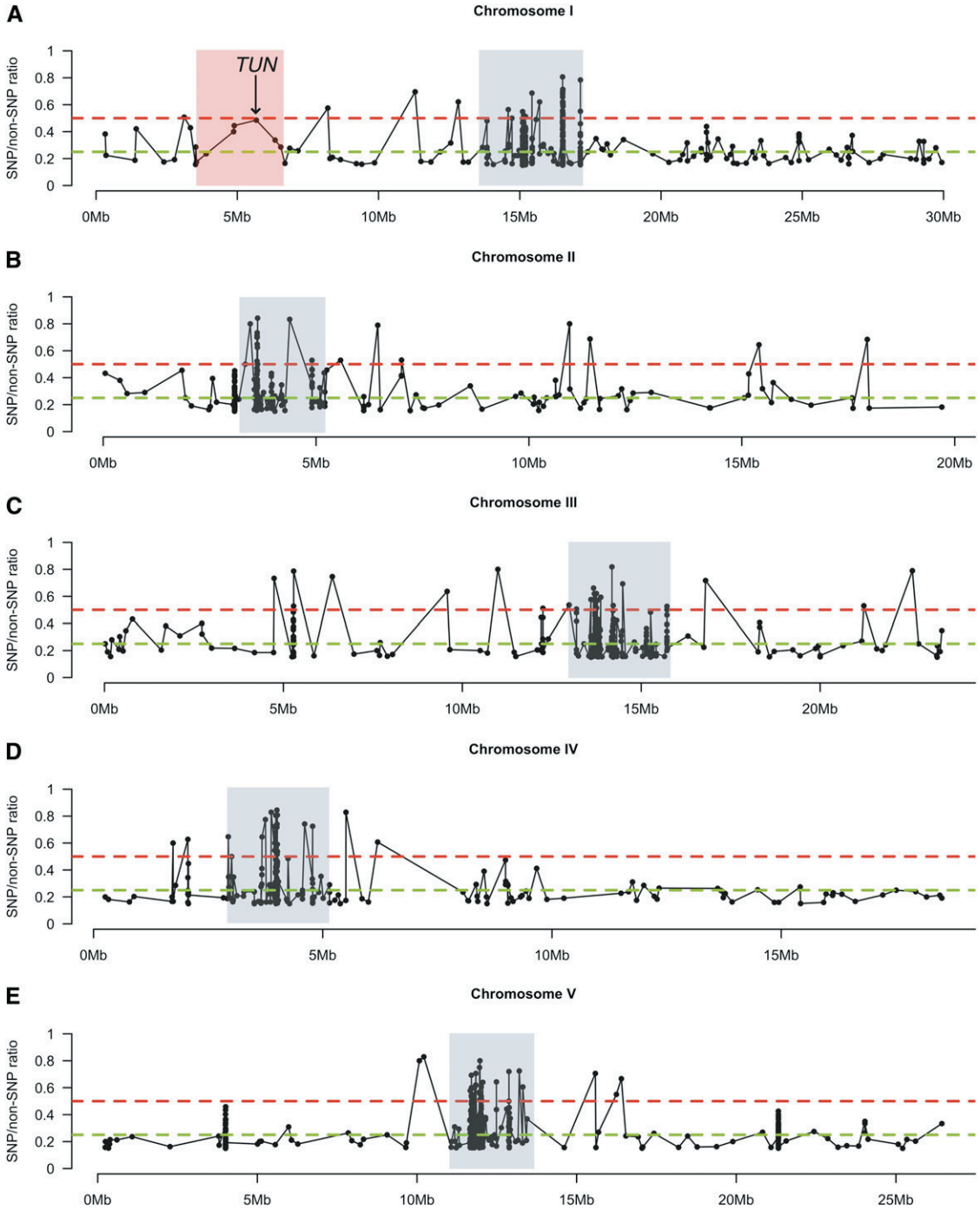


Figure 3 SNP/non-SNP ratio plots. The SNP/non-SNP ratio of all heterozygous SNPs is calculated and plotted against the chromosomal position of the heterozygous SNPs. The red dashed line marks the SNP/non-SNP ratio at 0.5, where the causative SNP should be; the green dashed line marks the SNP/non-SNP ratio at 0.25, where all other SNPs should locate. The red shading marks the genetically linked and selected region on chromosome I with the causative SNP in *At1g16570* (arrow). The gray shading marks the centromeric regions with a high SNP density, likely due to a poor mapping quality in these regions. (A–E) Chromosome I, II, III, IV, and V, respectively.

events close to the mutation, additional enhancer/suppressor mutations in the vicinity of the original mutation cannot be mapped. Furthermore, second-site modifier screens are often performed in complex, tailor-made backgrounds involving several mutants and/or transgenes (Page and Grossniklaus 2002). It is very hard to generate the identical genetic/transgenic constitution in two distinct accessions. By using SRM, the enhancer/suppressor mutant has to be backcrossed only to the original mutant background, no matter how complex it is.

Finally, SRM can be applied to any genetic system. In fact, we expect that SRM can also be applied in organisms without a well-annotated genome. As mentioned above, plotting the

SNP/non-SNP ratio over the chromosomal positions and visually identifying flat curves indicating a region under selection can be statistically tested (File S1 and File S2) and used to identify the causative SNP. This is also possible with partly assembled and poorly annotated genomes.

In conclusion, we successfully identified a gene disrupted in a gametophyte lethal mutant in *A. thaliana* by SRM. The need for relatively few individuals and only two rounds of backcrosses, which are needed in any case to purify the genetic background after mutagenesis, make this a very versatile and useful method for any genetic organism.

Acknowledgment

We thank A. Patrigiani for preparing the library for SOLiD sequencing and R. Schlapbach for access to the facilities of the Functional Genomics Center Zürich. We also thank S. A. Kessler and J. Jaenisch for helpful comments on the manuscript. This work was supported by grants from the European Research Council and the Swiss National Science Foundation to U.G. The costs for next-generation sequencing were covered by a project of the University Research Priority Program in Functional Genomics/System Biology of the University of Zürich.

Note added in proof: While our paper was under review, Abe *et al.* (2012) published a similar method based on the distinct segregation ratios of linked and unlinked SNPs to map homozygous mutants in rice. This shows that SRM can be applied to map homozygous mutants, as we suggested in our manuscript.

Literature Cited

Abe, A., S. Kosugi, K. Yoshida, S. Natsume, H. Takagi *et al.*, 2012 Genome sequencing reveals agronomically important loci in rice using MutMap. Nat. Biotechnol. 30: 174–178.

Austin, R. S., D. Vidaurre, G. Stamatiou, R. Breit, N. J. Provart *et al.*, 2011 Next-generation mapping of *Arabidopsis* genes. Plant J. 67: 715–725.

Blumenstiel, J. P., A. C. Noll, J. A. Griffiths, A. G. Perera, K. N. Walton *et al.*, 2009 Identification of EMS-induced mutations in *Drosophila melanogaster* by whole-genome sequencing. Genetics 182: 25–32.

Irvine, D. V., D. B. Goto, M. W. Vaughn, Y. Nakaseko, W. R. McCombie *et al.*, 2009 Mapping epigenetic mutations in fission yeast using whole-genome next-generation sequencing. Genome Res. 19: 1077–1083.

Jander, G., S. R. Norris, S. D. Rounsley, D. F. Bush, I. M. Levin *et al.*, 2002 *Arabidopsis* map-based cloning in the post-genome era. Plant Physiol. 129: 440–450.

Lukowitz, W., C. S. Gillmor, and W. R. Scheible, 2000 Positional cloning in *Arabidopsis*: Why it feels good to have a genome initiative working for you. Plant Physiol. 123: 795–805.

Page, D., and U. Grossniklaus, 2002 The art and design of genetic screens: *Arabidopsis thaliana*. Nat. Rev. Genet. 3: 124–136.

Sarin, S., S. Prabhu, M. M. O'Meara, I. Pe'er, and O. Hobert, 2008 *Caenorhabditis elegans* mutant allele identification by whole-genome sequencing. Nat. Methods 5: 865–867.

Schneeberger, K., S. Ossowski, C. Lanz, T. Juul, A. H. Petersen *et al.*, 2009 SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat. Methods 6: 550–551.

Sega, G. A., 1984 A review of the genetic effects of ethyl methanesulfonate. Mutat. Res. 134: 113–142.

Smith, D. R., A. R. Quinlan, H. E. Peckham, K. Makowsky, W. Tao *et al.*, 2008 Rapid whole-genome mutational profiling using next-generation sequencing technologies. Genome Res. 18: 1638–1642.

Srivatsan, A., Y. Han, J. Peng, A. K. Tehranchi, R. Gibbs *et al.*, 2008 High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. PLoS Genet. 4: e1000139.

Till, B. J., C. Burtner, L. Comai, and S. Henikoff, 2004 Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res. 32: 2632–2641.

Zuryn, S., S. Le Gras, K. Jamet, and S. Jarriault, 2010 A strategy for direct mapping and identification of mutations by whole-genome sequencing. Genetics 186: 427–430.

Communicating editor: C. D. Jones

PART 2

MUM OR DAD - WHO IS IN CHARGE?

PARENTAL EFFECTS DURING EMBRYOGENESIS IN *ARABIDOPSIS THALIANA*

PART 2
ABSTRACT

To study parental effects in young *Arabidopsis* embryos, we first developed a method allowing efficient and rapid isolation of early-stage embryos from *Arabidopsis thaliana* seeds (Chapter 2.1; Raissig et al., *in press*). This method enabled the isolation of young hybrid embryos to produce an allele-specific embryonic transcriptome. The analysis of parental contribution to the young embryonic transcriptome revealed a strong bias towards maternal-specific transcripts, supporting genetic data that suggested maternal control of paternal gene activity in early embryogenesis (Autran et al., 2011, Chapter 2.2.1). The genetic data was produced by our lab during the master thesis of Michael T. Raissig and by Daphné Autran and Daniel Grimanelli at the IRD in Montpellier, France. In contrast, a recent publication in *Nature* reports equal parental contribution during early embryogenesis in *Arabidopsis* and suggests that the maternal bias reported by Autran and colleagues (2011) results from maternal sporophytic contamination and that the observed bias is not genuine (Nodine and Bartel, 2012). The obvious discrepancy between the two articles is discussed in Chapter 2.2.2. However, both articles independently report genes that show parent-of-origin-dependent monoallelic expression in early *Arabidopsis* embryos. We analyzed the dataset of Autran and colleagues (2011), called 80 potentially imprinted genes and confirmed 12 genes to be likely regulated by genomic imprinting in the embryo (Chapter 2.3; Raissig et al., *in preparation*). Monoallelic gene expression in the *Arabidopsis* embryo is partly regulated by PRC2 and the imprint is erased during late embryogenesis or early vegetative development (Chapter 2.3).

Importantly, each subchapter is built like a publication (even if it is not published or submitted yet, like Chapter 2.3) containing introduction, material and methods, results, discussion and separate references.

PART 2
CHAPTER 1

EFFICIENT AND RAPID ISOLATION OF EARLY-STAGE EMBRYOS FROM *ARABIDOPSIS THALIANA* SEEDS

NOTE

All of CHAPTER 2.1 is *in press* in the *Journal of Visualized Experiments* (*JoVE*). Contributing authors are Michael T. Raissig, Valeria Gagliardini, Johan Jaenisch, Ueli Grossniklaus and Célia Baroux.

MTR as first author and CB as communicating author developed the method together and VG helped to improve it and added to its transferability between different users and labs. JJ established the fluorescence *in situ* hybridization (FISH) protocol on whole-mount embryos. Since the method is not yet printed we put the accepted manuscript version in this thesis.

SHORT ABSTRACT:

We report an efficient and simple method to isolate embryos at early stages of development from *Arabidopsis thaliana* seeds. Up to 40 embryos can be isolated in 1h to 4h, depending on the downstream application. The procedure is suitable for transcriptome, DNA-methylation, reporter gene expression, immunostaining and fluorescence *in-situ* hybridization analyses.

LONG ABSTRACT:

In flowering plants, the embryo develops within a nourishing tissue – the endosperm – surrounded by the maternal seed integuments (or seed coat). As a consequence, the isolation of plant embryos at early stages (1-cell to globular stage) is technically challenging due to their relative inaccessibility. Efficient manual dissection at early stages is strongly impaired by the small size of young *Arabidopsis* seeds and the adhesiveness of the embryo to the surrounding tissues. Here, we describe a method that allows the efficient isolation of young *Arabidopsis* embryos, yielding up to 40 embryos in 1h to 4h, depending on the downstream application. Embryos are released into isolation buffer by slightly crushing 250-750 seeds with a plastic pestle in an Eppendorf tube. A glass microcapillary attached to either a standard laboratory pipette (via a rubber tube) or a hydraulically controlled microinjector is used to collect embryos from droplets placed on a multi-well slide on an inverted light microscope. The technical skills required are simple and easily transferable, and the basic setup does not require costly equipment. Collected embryos are suitable for a variety of downstream applications such as RT-PCR, RNA sequencing, DNA methylation analyses, fluorescence *in-situ* hybridization (FISH), immunostaining**,** and reporter gene assays.

INTRODUCTION

The embryo of flowering plants is surrounded by the endosperm, a nutritive tissue derived from a second fertilization event. Both embryo and endosperm are surrounded by several cell layers of the seed coat. Collectively these tissues form a seed, which develop inside the fruit. Thus, tissue- and cell-specific analyses of *Arabidopsis* embryos are strongly impaired due their inaccessibility. Nevertheless, embryos at the late-globular or later stages are relatively well amenable to manual dissection by using fine tungsten needles under the stereomicroscope, or by applying slight pressure on the seed using forceps to extract them. Such techniques were successfully used for transcriptome or epigenome profiling analyses such as microarray hybridization, bisulfite sequencing, or RNA sequencing (eg. Gehring et al., 2009; Gehring et al., 2011; Müller & Sheen, 2008). In contrast, studies of embryos at the zygote to early globular stage remain technically challenging. To date, only a few studies have reported transcriptome analyses on young embryos using either laser-capture microdissection (LCM) of embryonic tissues from fixed seed sections (Harada-Goldberg Laboratories, GSE12404) or manual extraction of individual embryos from within seeds using fine tools (Xiang et al., 2011). However, LCM equipment is not commonly available and manual embryo extraction at early stages is time consuming and requiring excellent dissection skills that are not easily transferable. In addition to genome-wide analyses, *in situ* gene expression analyses are also difficult to perform on young, whole-mount embryos of *Arabidopsis*. To some extent, young embryos can be released on microscope slides by gentle pressure on the seeds and used for reporter gene assays or protein detection by immunostaining (for example see Baroux et al., 2007; Nawy et al., 2010). This technique, however, does not allow high-throughput embryo isolation, thus hindering quantitative analyses.

Therefore, we developed an efficient and rapid protocol for early embryo isolation from *Arabidopsis* seeds that is simple to set up, easily transferable, and suitable for a variety of downstream applications. The basic principle is to gently crush seeds – dissected from young siliques in an Eppendorf tube using a plastic pestle in an appropriate isolation buffer. The seed extract is placed in droplets on a multi-well slide and is screened for the presence of released embryos at the desired stage using an inverted microscope. Embryos are collected using a glass microcapillary attached to a microinjector or a standard laboratory pipette. For molecular applications, embryos are washed twice by repeated release into drops of new isolation buffer before transferring them to the destination buffer in a minimal volume. For cytological applications (reporter assays, immunostaining, FISH), washing steps can be omitted.

The method offers several advantages: (i) it yields 25-40 embryos in ca 45 min for cytological applications or in 3-4h for molecular applications (including the washing steps), (ii) it allows isolation of specific embryonic stages, (iii) it is easily transferable to other persons and laboratories due to its simple setup, (iv) it requires affordable equipment for the basic setup which is amenable to upgrades, and (v) it was successfully used for various downstream applications such as RNA sequencing (Autran et al., 2011), gene-specific DNA-methylation analysis (Wöhrmann et al., 2012), reporter assays (Raissig et al., 2011; and Raissig et al., in prep.), and FISH (J. Jaenisch, U. Grossniklaus, C. Baroux unpublished, see Figure 2-1-5)

PROCEDURE

The procedure is summarized in the flowchart shown in Figure 2-1-1. The microcapillaries and the instrumental setup are shown in Figure 2-1-2 and Figure 2-1-3, and typical steps of embryo isolation are shown in Figure 2-1-4.

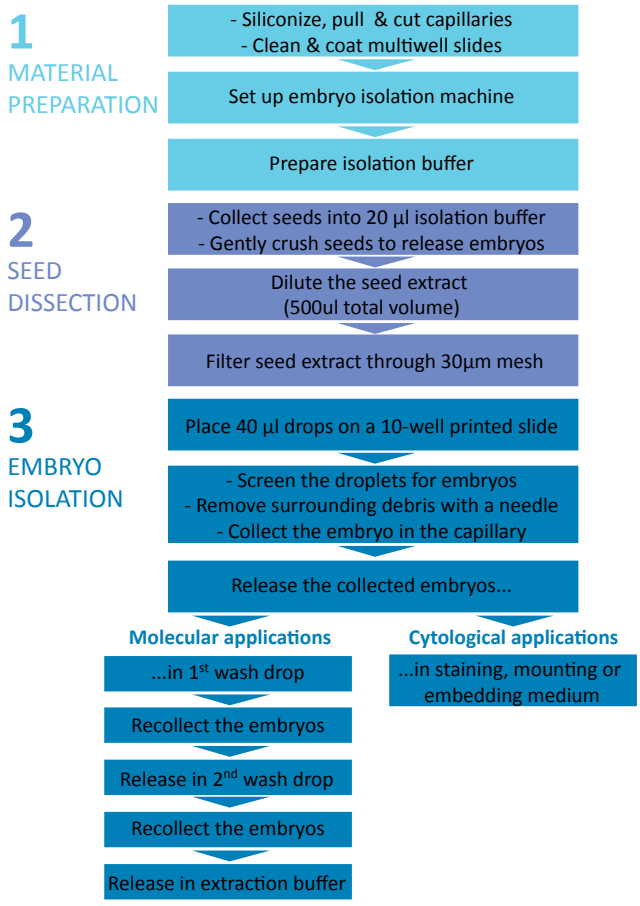


Figure 2-1-1: Flow chart of the embryo isolation procedure. The protocol is divided in three parts: 1 – Material preparation; 2 – Seed dissection; 3 – Embryo isolation

1.) Material and buffer preparation

1.1) Silicon coating of glass microcapillaries: Place the microcapillaries in a 15mL Falcon tube with ~5mL of Sigmacote (Sigma) and invert several times. Remove the solution, place the falcon tube containing the capillaries in an aluminum foil and bake them for 3 hours at 60°C. Store at room temperature.

1.2) Obtain ~50-100 µm-diameter microcapillary tips: Pull 1mm-diameter glass capillaries either manually over a Bunsen burner or by using a commercial puller (vertical filament puller or micropipet puller); use a diamond-tip pen or blade to cut the tip of the pulled capillary to create the desired opening. Select the best-shaped capillaries under a stereo microscope. The opening should be 50-100 µm (Figure 2-1-2).

1.3) Slide preparation:

1.3.1. Siliconize clean slides by covering all the wells with Sigmacote (ca 1mL/ slide) for 5 min, remove. Bake 3 hrs in aluminium foil. Store at room temperature.

1.3.2. Wash the multi-well glass microscopic slides for 10 min in 10% SDS, 2x 2 min in nuclease-free water (autoclaved DEPC-ddH20), 2 min in 70% ethanol, 2 min in 100% ethanol, air-dry. All steps are done in autoclaved Coplin jars. Slides can be re-used multiple times providing thorough cleaning between each usage.

1.3.3. Just prior to embryo isolation spread ca. 0.5µL of 10mg/mL bovine serum albumine (BSA) with a pipette over the whole surface of each well and air-dry.

1.4) Microscope and capillary setup:

1.4.1. Use an inverted microscope with a 10x and 20x magnification objective. Optimize the light contrast (embryos are quite transparent).

1.4.2. Place a micromanipulator to hold the glass capillary beside the microscope. The glass capillary is connected to a microinjector (Figure 2-1-3A) or to a

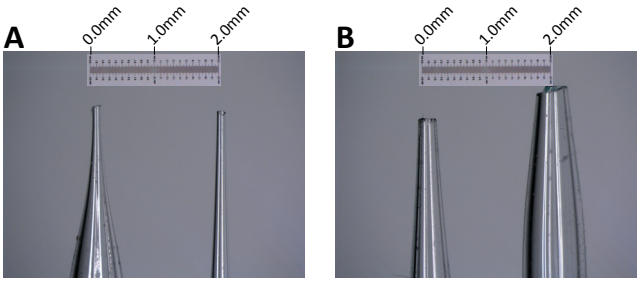


Figure 2-1-2: Microcapillary tips. A. High-quality microcapillary tips with a smooth opening of around 100 µm. B. Low-quality microcapillary tips with wider opening and irregular contour (those tips are acceptable for cytological applications only). Scale bar: 2 mm.

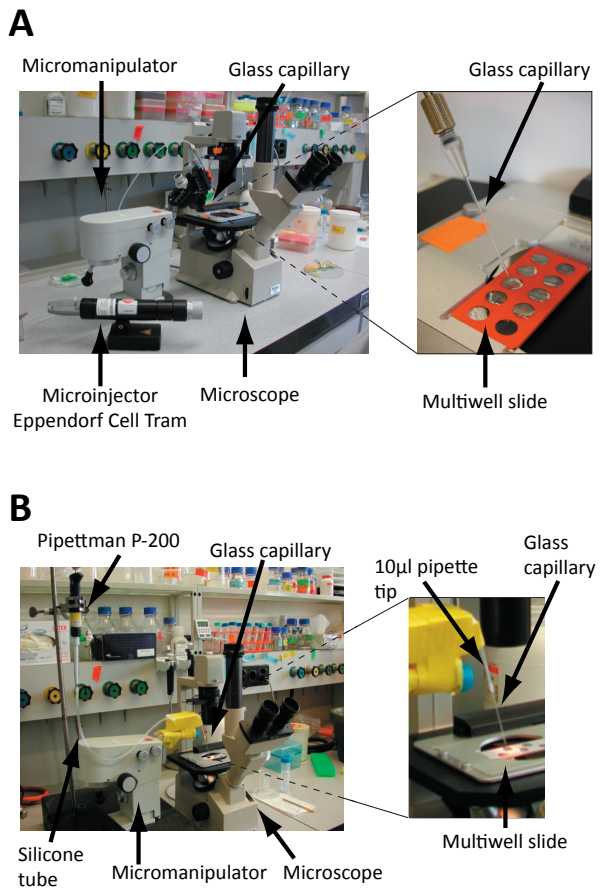


Figure 2-1-3: Set up of the embryo isolation microscope. A. and B. Screening is performed on microscope slides under an inverted microscope. Embryos are collected using a glass microcapillary fixed on a micromanipulator to precisely control its position (see also Figure 4) and connected to either a microinjector (A) or standard laboratory pipette (B). A. The glass microcapillary is linked to a hydraulically controlled microinjector (eg. Eppendorf Cell Tram Vario) for a precise control during embryo collection B. The glass microcapillary is attached to a standard P-200 pipette via a rubber flex tube. Embryo collection and release is controlled by turning the calibration wheel of the pipette. Cut-end pipette tips are used as connectors and the junctions are sealed with parafilm. The microcapillary is fixed on the micromanipulator via polystyrene blocks, Falcon tubes and tape.

regular P-200 pipette via a rubber tube (Figure 2-1-3B, see Discussion for a detailed description of this setup).

1.4.3. Place the capillary above the microscope slide (*ca* 70°, Figure 2-1-3) and adjust the position to have the opening in the field of view.

1.5) Buffers:

Table 1 lists the isolation and destination buffers depending on the downstream applications.

1.5.1. Prepare ~1mL isolation buffer per sample freshly before use and keep it on ice.

1.5.2. Prepare the destination buffer in a 0.5mL Eppendorf low-binding tube on ice (molecular applications) or on a microscope slide (cytological applications) in a humid chamber.

2.) Seed dissection and embryo extraction

2.1) Synchronisation of seed development:

Emasculate flowers and keep them 2 days in the growth chamber while avoiding contact of the exposed pistils with other flowers, then pollinate them (e.g. Rea et al., 2011). Test the stage of development under your growth conditions by microscopic investigation of cleared seeds. With our growth conditions (16h light at 21°C, 8h dark at 18°C and 70% humidity) seeds collected 2.5 days after pollination (DAP) yielded mainly 2-4 cell embryos and seeds collected 3.5 – 4 DAP yielded globular embryos.

2.2) Seed dissection and rupture

2.2.1. Remove the seeds from 15 siliques (~2.5 DAP) under a stereomicroscope with forceps and insulin needles.

2.2.2. Immerse the seeds in 20µL isolation buffer in a 2mL round-bottom Eppendorf tube placed on ice.

2.2.3. Gently crush the seeds with a plastic pestle (pre-cleaned with 10% SDS, rinsed with DEPC-ddH20 and washed with 70% Ethanol) to release the embryos until the seed extract is cloudy (*ca* 30 sec). The force to apply is to be determined by every user upon trial.

2.2.4. Rinse the pestle with 300 µL of isolation buffer to wash the pestle and dilute the sample.

2.2.5. Spin-down the extract at 5g for 5 sec. Gently resuspend the pelleted extract by pipetting up-and-down 2-3x

using a Pipetman P-200.

2.2.6. Filter the extract with a 30µm nylon mesh (mounted on tube adaptors, e.g. from Partec Celltricks). Rinse the mesh with an additional 200µL isolation buffer.

3.) Embryo isolation

3.1) Slide preparation

3.1.1. Place a clean, BSA-coated and siliconized multi-well slide on the stage of the inverted microscope, resuspend the filtered seed extract by pipetting gently up and down and pipette 2 droplets of 40-50 µL seed extract into 1 or 2 wells. Screening only 1 or 2 drops at a time prevents evaporation of the sample.

3.1.2. Place 50 µL of fresh isolation buffer (1st wash drop) in a well of a different slide prepared as before. Keep this slide in a covered, humid chamber to prevent evaporation.

3.2) Screen, clean, collect.

3.2.1. Screen the droplets of seed extract for embryos at the desired stage with the 10x magnification objective. If necessary, confirm the stage with the 20x magnification. The embryos usually sink to the bottom of the slide.

3.2.2. Manually remove debris around the embryo with a tungsten needle, an insulin needle or similar equipment.

3.2.3. Move the glass capillary near the embryo using the micromanipulator, take up the embryo with as little solution as possible

3.2.4. Collect several embryos (e.g. all of one droplet) and release them in the 1st wash drop (molecular application) or in destination buffer (cytological applications). Each collecting round should be kept within 5-10min and embryos should be collected in a minimal volume (<1-5µL).

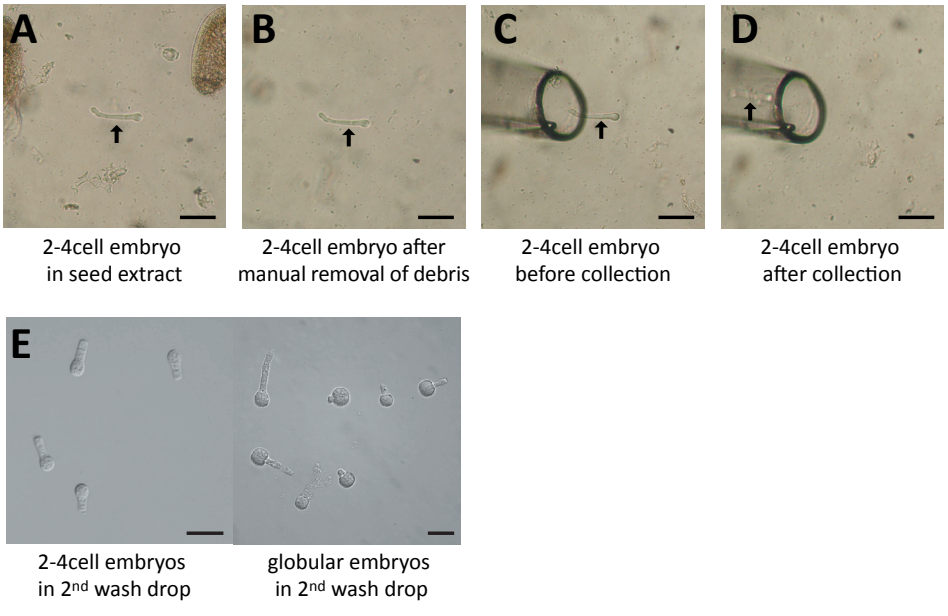


Figure 2-1-4: Embryo isolation process. A. A 2-4 cell embryo (arrow) was identified in the seed extract (screening droplet) and is surrounded by debris (seed coat fragments). B. The surrounding debris was removed manually using a needle. C. The glass capillary was moved right beside the embryo (arrow). D. The embryo was collected and is now within the capillary (arrow). E. Several embryos in the last drop following washes. Scale = 50 µm.

3.2.5. Repeat the screening and collection until the desired amount of embryos is gathered in the 1st wash drop (centrally, if possible, to facilitate recollection).

3.2.6. Recollect all embryos at once from the wash drop (if debris are carried over, remove them with a needle before recollection).

3.2.7. Release the embryos into a 2nd wash drop of 50 µL. Repeat 3.2.6.

3.2.8. Release the embryos in the destination buffer. The transfer should involve only a contact, and not immersion, of the capillary tip.

3.2.9. Replace the microcapillary for the next sample.

REPRESENTATIVE RESULTS

Our embryo isolation procedure (Figure 2-1-1) allows isolation of up to 40 embryos in 4 hours if washes are performed, e.g. for molecular applications, or in less than an hour if washes are omitted, e.g. for cytological applications. Figure 2-1-2 displays high and low quality microcapillary tips and Figure 2-1-3 shows the setup of the embryo isolation machine. Figure 2-1-4 displays the process of embryo isolation on the inverted microscope.

We successfully applied our procedure for various applications published in several recent articles. The method was originally developed to analyze the parental contribution to the early embryonic transcriptome. Hybrid embryos generated through crosses between two different *Arabidopsis* accessions (Landsberg *erecta* (L*er*) and Columbia (Col-0)) were isolated at the 2-4 cell stage. Total RNA was extracted, cDNA libraries were produced using linear amplification and sequenced on a SOLiD platform. The allele-specific transcriptomes were generated based on SNP analysis (Autran et al., 2011). To control our embryonic cDNA libraries prior to sequencing we amplified embryo-specific transcripts, *WUSCHEL-HOMEOBOX 2, 8* and *9 (WOX2, WOX8, WOX9)* and *ACTIN 11* (Breuninger et al., 2008; Köhler et al., 2005; Figure 2-1-5A).

Additionally, this method was used to isolate young embryos to analyze the embryonic DNA-methylation patterns at specific loci in the genome. Embryos were isolated and washed in 1x TE-buffer. Small-scale bisulfite sequencing was performed as described (Wöhrmann et al., 2012).

Similarly, embryo isolation has proved very useful in reporter gene assays with low embryonic expression, and where the maternal seed coat confounds detection or is masked by reporter expression in tissues surrounding the embryo (endosperm, seed coat). Embryos carrying the reporter transgene are stained (ß-glucuronidase; GUS) or directly analyzed (Green or Red Florescent Protein; GFP, RFP) following isolation. An example is given in Figure 2-1-5B for embryos expressing a GUS reporter gene under the control of the *MEDEA* promoter (p*MEA;* Raissig et al., 2011). The relatively ease with which many embryos can be isolated also allows for quantitative analyses (e.g. number of embryos stained in different genetic backgrounds, Raissig, Grossniklaus et al. in preparation).

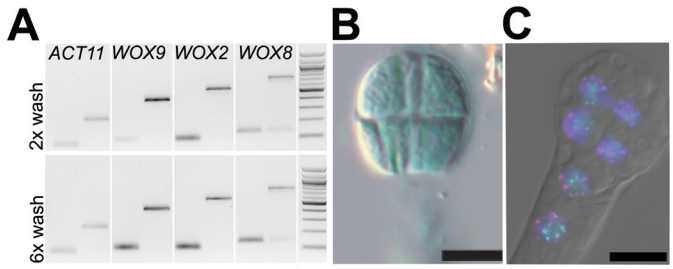


Figure 2-1-5: Downstream applications of embryo isolation. A. Gene expression analyses: PCR amplification of *WOX9*, *WOX2*, *WOX8* and *ACTIN11* on cDNA libraries from 2-4 cell embryos washed 1x times (1st lane) and on genomic DNA (2nd lane). B. Reporter assay: An 8-cell embryo isolated from plants carrying a p*MEA::GUS* construct was stained on a slide in a standard GUS staining solution (reproduced from Raissig et al., 2011 after permission from *The Plant Cell*, ASPB copyright). C. Fluorescent *In-Situ* Hybridization (FISH): an 8-cell embryo was hybridized with probes against centromeric repeats (red), and 45S rDNA repeats (green) before indirect immunodetection. The dual-color FISH images were overlaid with the DAPI counterstaining and DIC images (DM6000B epifluorescence microscope, Leica, Germany)

Finally, we successfully applied Fluorescent *In Situ* Hybridization (FISH) and immunostaining techniques to study the nuclear architecture in isolated embryos embedded in acrylamide pads on slides. An example of FISH using probes against the centromere repeats and nucleolar organizing regions is shown in Figure 2-1-5C.

DISCUSSION

We developed an embryo isolation protocol that is rapid, effective, and can be easily transferred to other laboratories.

The equipment described here consists of an inverted microscope, a micromanipulator, glass microcapillaries, a vertical filament puller and a microinjector (Figure 2-1-3A). The setup is similar to the one described for single animal cell isolation for transcriptomics analyses (Morris et al., 2011). We also successfully worked with a more basic setup where glass microcapillaries were manually stretched over a flame (and cut with a diamond-blade) and operated via a standard laboratory pipette (Pipetman P-200) instead of a microinjector (Figure 2-1-3B). In that case the microcapillary was attached to the pipette via a rubber tube. 10 µL filtertips were used to make the junctions, which were wrapped with parafilm to make them air-tight. The capillary - held by the pipette tip - was fixed on the micromanipulator arm using polystyrene blocks and tape (Figure 2-1-3B). This basic setup proved to be efficient and reliable (Autran et al., 2011). Nevertheless, one of the main difficulties was the maintening of air-tight junctions between the microcapillary and the pipette, especially when changing the capillary. Verifying and adjusting the pressure in the capillary – to ensure a fine-tuned embryo collection in a minimal volume - can be time-consuming using this basic setup. The improved setup with a hydraulically controlled micromanipulator and a vertical filament puller is a considerable improvement and saves time.

The skills required for the successful application of this method are easily transferable. Five users in our laboratory successfully learnt the method in a relatively short time. Some simplifications of the protocol are possible depending on the user's ease in manipulation. For instance, it is possible to skip the steps 2.2.5 and 2.2.6 while dissecting only 5 siliques (instead of 15) and cleaning the embryo surroundings thoroughly from debris with a tungsten needle (this procedure was applied in Autran et al., 2011). Synchronisation of seed development through emasculation and delayed pollination is also facultative but recommended to increase the number of embryos at the same developmental stage, particularly at early stages. In addition, embryo isolation can be expedited if washing steps are omitted, e.g. for cytological applications. Furthermore, slide siliconization (step 1.3.2) is not necessary for users working quickly, such that embryos which stick to the glass surface over time is not a problem.

Whether filtering or manual cleaning is applied, it is extremely important to avoid carry-over of tissue debris that creates potential contamination for downstream RNA or DNA applications. This issue was raised recently following two transcriptome profiling studies on isolated *Arabidopsis* embryos with different outcome (Autran et al., 2011; Nodine and Bartel, 2012). While we described a strong maternal dominance at early stages (88% sequencing reads were of maternal origin; Autran et al., 2011), the latest study report on equal parental contribution (Nodine and Bartel, 2012). In fact, the two studies differ in their experimental setup including the mode of embryo isolation for transcriptome analyses. By contrast to the bulk embryo extraction protocol described here, the authors of the latest study dissected the embryos manually from within individual seeds, a lengthy procedure requiring excellent dissection skills. This procedure apparently required 2 or more washes to avoid carry-over of maternal information in the embryo transcriptome. Contaminant cells may be expected in manual extraction given the small size and relative inaccessibility of the embryos to dissection needles. By contrast, our procedure allows (i) the dilution of the isolated embryos and surrounding debris in a large volume (500 µL) before embryo collection, (ii) the visual selection of embryos devoid of adhesive, contaminant tissues and (iii) the dilution of possible RNA contaminants -issued from ruptured cells- through washing. While it is possible to consider only one wash if the embryos were visually very clean and collected in less than 5µL (Autran et al., 2011), a second wash step is recommended, especially for first time users that may collect the embryos in several rounds (i.e. with additive volume). Embryos should be collected with as little solution as possible, allowing

maximal dilution of potential contaminants in every washing drop. Each collecting round should be kept short (within 5-10min) to allow maximum recovery upon release in the wash drop (prolonged stay in the capillary tends to reduce the recovery rate). Furthermore, the transfer of the collected embryos into the extraction medium (following the wash steps) should only involve a contact with the opening of the microcapillary tip and not immersion of the tip, as this might as well transfer debris sticking on the microcapillary wall above the opening. Finally, the capillary should be changed between each new seed extract"

Our protocol allows for several downstream applications by simply using an appropriate isolation buffer that preserve molecular integrity of RNA, DNA, chromatin, enzyme or fluorescent reporter. We successfully isolated embryos in RNA-protective isolation buffer for RNA extraction and transcriptomics analyses (Autran et al., 2011), 1x TE-buffer for DNA methylation analyses (Wöhrmann et al., 2012), 100mM Phosphate buffered saline (PBS) for FISH and immunostaining (Jaenisch, Grossniklaus and Baroux, unpublished, Figure 2-1-5C), and staining solution without substrate for GUS reporter assays (Jefferson et al., 1987; Figure 2-1-5B). The application most sensitive to isolation conditions/embryo quality is certainly RNA extraction and amplification. The quantity of extracted RNA from isolated embryos was rather low. From ca. 30 embryos at the 2-4 cell stage (thus 60-120 cells in total) we estimated an amount of ca 1ng of total RNA based on both Agilent 2100 Bioanalyzer (Agilent Technologies) and QUBiT (Invitrogen) measurements. Following RNA extraction, RNA quality was verified on an Agilent RNA 6000 Pico Chip (Agilent Technologies). However, the low amount of RNA extracted is not always sufficient to produce a reliable profile. Thus, further tests must be done on the amplified material (cDNA) (e.g. PCR detecting low-expressed, embryo-specific genes and/or a Bioanalyzer profile).

Finally, the ability to isolate embryos by visual criteria in our protocol is a great asset. Embryos from the same silique (*Arabidopsis* fruit) do not develop synchronously. Thus, working on whole silique extracts (e.g. for RT-PCR analyses or quantitative reporter assays) does not allow a perfect correlation with a given developmental stage. Isolating embryos at a specific developmental stage solves this problem. In addition, a similar problem presents when working with heterozygous mutants where siliques contain different embryo genotypes. Isolating embryos based on visual criteria to distinguish e.g. wild-type from mutant embryo phenotypes allows correlating downstream analyses with the genotype. We have done this successfully to analyze DNA methylation at specific loci in different genotypes (Schmidt, Raissig, Grossniklaus et al, in preparation).

ACKNOWLEDGMENTS:

We would like to thank Tal Nawy and Martin Bayer for their advice on embryo isolation. MTR, VG, UG and CB devised the embryo isolation equipment. MTR, VG and CB developed the embryo isolation protocol. MTR, VG and CB established the protocol, isolated the embryos, and generated embryo cDNA, VG performed the PCR, MTR the GUS staining, JJ the FISH experiments. MTR, VG, CG and UG wrote the manuscript. This work was funded by the University of Zürich, a Fellowship of the Roche Research Foundation (to MTR), and grants from the Swiss National Foundation (to UG and CB).

References:

Autran, D., Baroux, C., Raissig, M.T., Lenormand, T., Wittig, M., Grob, S., Steimer, A., Barann, M., Klostermeier, U.C., Leblanc, O., Vielle-Calzada, J.-P., Rosenstiel, P., Grimanelli, D., and Grossniklaus, U. (2011). Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. Cell **145**: 707–719.

Baroux, C., Pecinka, A., Fuchs, J., Schubert, I., and Grossniklaus, U. (2007). The triploid endosperm genome of *Arabidopsis* adopts a peculiar, parental-dosage-dependent chromatin organization. Plant Cell **19**: 1782–94.

Breuninger, H., Rikirsch, E., Hermann, M., Ueda, M., and Laux, T. (2008). Differential expression of *WOX* genes mediates apical-basal axis formation in the *Arabidopsis* embryo. Dev Cell **14**: 867–76.

Gehring, M., Bubb, K.L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science **324**: 1447–1451.

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic Analysis of Parent-of-Origin Allelic Expression in *Arabidopsis thaliana* Seeds. PLoS ONE **6**: e23687.

Harada-Goldberg Laboratories Expression data in seed tissues from the Harada-Goldberg *Arabidopsis* LCM GeneChip Data Set, Seedgenenetwork at http://estdb.biology.ucla.edu/ seed/, Gene Expression Omnibus accession series GSE12404.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J **6**: 3901–7.

Köhler, C., Page, D.R., Gagliardini, V., and Grossniklaus, U. (2005). The *Arabidopsis thaliana MEDEA Polycomb* group protein controls expression of *PHERES1* by parental imprinting. Nat Genet **37**: 28–30.

Morris, J., Singh, J.M., and Eberwine, J.H. (2011). Transcriptome analysis of single cells. J Vis Exp **25**: 2634.

Müller, B. and Sheen, J. (2008). Cytokinin and auxin interaction in root stem-cell specification during early embryogenesis. Nature **453**: 1094–7.

Nawy, T., Bayer, M., Mravec, J., Friml, J., Birnbaum, K.D., and Lukowitz, W. (2010). The GATA factor *HANABA TARANU* is required to position the proembryo boundary in the early *Arabidopsis* embryo. Dev Cell **19**: 103–13.

Nodine, M.D. and Bartel, D.P. (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature **482**: 94–97.

Raissig, M.T., Baroux, C., and Grossniklaus, U. (2011). Regulation and Flexibility of Genomic Imprinting during Seed Development. Plant Cell **23**: 16–26.

Rea, M., Chen, M., Luan, S., Bhangu, D., Braud, M., and Xiao, W. (2011). Determination of DNA methylation of imprinted genes in *Arabidopsis* endosperm. J Vis Exp **28**: 2327

Wöhrmann, H.J., Gagliardini, V., Raissig, M.T., Wehrle, W., Arand, J., Schmidt, A., Tierling, S., Page, D.R., Schöb, H., Walter, J., and Grossniklaus, U. (2012). Identification of a DNA methylation-independent imprinting control region at the *Arabidopsis MEDEA* locus. Genes Dev **26**: 1837–1850.

Xiang, D., Venglat, P., Tibiche, C., Yang, H., Risseeuw, E., Cao, Y., Babic, V., Cloutier, M., Keller, W., Wang, E., Selvaraj, G., and Datla, R. (2011). Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*. Plant Phys **156**: 346–356.

PART 2
CHAPTER 2

2.2.1: MATERNAL EPIGENETIC PATHWAYS CONTROL PARENTAL CONTRIBUTIONS TO *ARABIDOPSIS* EARLY EMBRYOGENESIS

2.2.2: *ARABIDOPSIS* EMBRYOGENESIS - MATERNAL DOMINANCE OR EQUAL PARENTAL POLICY?

NOTE

All of CHAPTER 2.2.1 is published as Autran D*, Baroux C*, Raissig MT, et al. (2011) *Cell* **145:** 707 – 719.

**these authors contributed equally to this work*

MTR contributed to this work as follows: During MTR's master thesis entitled "Maternal Control of Paternal Genome Activation" (2007) he analyzed the activation and expression dynamics of paternal reporter lines, either in wild-type mother plants (Figure 3) or in mother plants mutant for CAF1 nucleosome assembly complex subunits, which are involved in the activation of paternal reporter genes (Figure 7). At the beginning of his PhD thesis, MTR isolated hybrid embryonic samples (wild type and mutant), extracted RNA and amplified cDNA libraries for RNA sequencing and allele-specific transcriptome analysis together with CB (Figure 1, 2, 4D-G). Finally, he aided in writing and editing the present manuscript and changed all figures according the style of *Cell*. Supplemental information to this publication can be found in Appendix A6.

In CHAPTER 2.2.2 a recent and controversial publication reporting equal parental contribution during early embryogenesis in *Arabidopsis* (Nodine M and Bartel DP (2012) *Nature* **482:** 94-97) is discussed and possible technical and biological aspects causing the discrepancy between the two reports are presented.

Cell

Maternal Epigenetic Pathways Control Parental Contributions to *Arabidopsis* Early Embryogenesis

Daphné Autran,[1,6] Célia Baroux,[2,6] Michael T. Raissig,[2] Thomas Lenormand,[3] Michael Wittig,[4] Stefan Grob,[2] Andrea Steimer,[2] Matthias Barann,[4] Ulrich C. Klostermeier,[4] Olivier Leblanc,[1] Jean-Philippe Vielle-Calzada,[5] Phillip Rosenstiel,[4] Daniel Grimanelli,[1,*] and Ueli Grossniklaus[2,*]

[1]Diversité, Adaptation et Développement des Plantes, Institut de Recherche pour le Développement, Université de Montpellier, UMR 232, Montpellier 34394, France

[2]Institute of Plant Biology and Zürich-Basel Plant Science Center, University of Zürich, Zürich CH-8008, Switzerland

[3]Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, Université de Montpellier, UMR 5175, Montpellier 34293, France

[4]Institute of Clinical Molecular Biology, Christian Albrechts University Kiel, Kiel 24105, Germany

[5]National Laboratory of Genomics for Biodiversity, CINVESTAV Guanajuato, Irapuato CP36500, Guanajuato, Mexico

[6]These authors contributed equally to this work

*Correspondence: daniel.grimanelli@ird.fr (D.G.), grossnik@botinst.uzh.ch (U.G.)

DOI 10.1016/j.cell.2011.04.014

SUMMARY

Defining the contributions and interactions of paternal and maternal genomes during embryo development is critical to understand the fundamental processes involved in hybrid vigor, hybrid sterility, and reproductive isolation. To determine the parental contributions and their regulation during *Arabidopsis* embryogenesis, we combined deep-sequencing-based RNA profiling and genetic analyses. At the 2–4 cell stage there is a strong, genome-wide dominance of maternal transcripts, although transcripts are contributed by both parental genomes. At the globular stage the relative paternal contribution is higher, largely due to a gradual activation of the paternal genome. We identified two antagonistic maternal pathways that control these parental contributions. Paternal alleles are initially downregulated by the chromatin siRNA pathway, linked to DNA and histone methylation, whereas transcriptional activation requires maternal activity of the histone chaperone complex CAF1. Our results define maternal epigenetic pathways controlling the parental contributions in plant embryos, which are distinct from those regulating genomic imprinting.

INTRODUCTION

In most animal species, the zygote is transcriptionally quiescent, and early embryogenesis is governed by maternal products stored in the oocyte prior to fertilization (Andéol, 1994). Depending on the species, zygotic genome activation (ZGA) takes place after one to several cell divisions. ZGA is a gradual process that relies on large-scale chromatin reprogramming leading to an increasing number of zygotically expressed genes (Tadros and Lipshitz, 2009). Maternal transcripts and proteins inherited from the gametes are progressively degraded, and biparental zygotic transcripts gradually take over the control of development. As a result, parental contributions to the embryonic transcriptome dynamically change during early development, with an initial maternal control that is of variable duration (1 to 15 cell cycles) (Baroux et al., 2008; Tadros and Lipshitz, 2009). Strikingly, such a maternal influence occurs in animals as evolutionarily divergent as insects, amphibians, and mammals. Understanding the parental contributions and the regulation of zygotic genome expression during early embryogenesis is a key question in developmental and evolutionary biology.

In flowering plants, the knowledge about the regulation and dynamics of parental contributions during early embryogenesis remains fragmented, despite its importance in understanding hybrid vigor, hybrid viability, parent-of-origin-dependent interploidy, and nonself pollination (xenia) effects determined by interactions of parental genomes after fertilization (Bushell et al., 2003; Jahnke et al., 2010; Meyer and Scholten, 2007; Pahlavani and Abolhasani, 2006). In flowering plants, double fertilization produces the zygote, which develops into the embryo (Movie S1 available online), and the endosperm, an embryo-nurturing tissue. Both fertilization products develop within maternal integuments, forming the seed. Genetic studies have shown that seed development is under maternal influence (Chaudhury and Berger, 2001), but the composite nature of the seed makes determining the origin of maternal effects complex. Recently, downregulation of RNA Polymerase II (PolII) in the mature *Arabidopsis* egg cell revealed that the embryo developed to the preglobular stage in absence of significant de novo transcription (Pillot et al., 2010b). Thus, de novo transcription of parental genomes is not an absolute requirement for early embryogenesis, but the timing, dynamics, and mechanisms of zygotic genome activation have yet to be elucidated. Reporter and profiling studies on whole seeds from *Arabidopsis* and maize have identified several transcripts with a dominant maternal representation at early stages, whereas paternal transcripts were detected only later (Baroux et al., 2001; Grimanelli et al.,

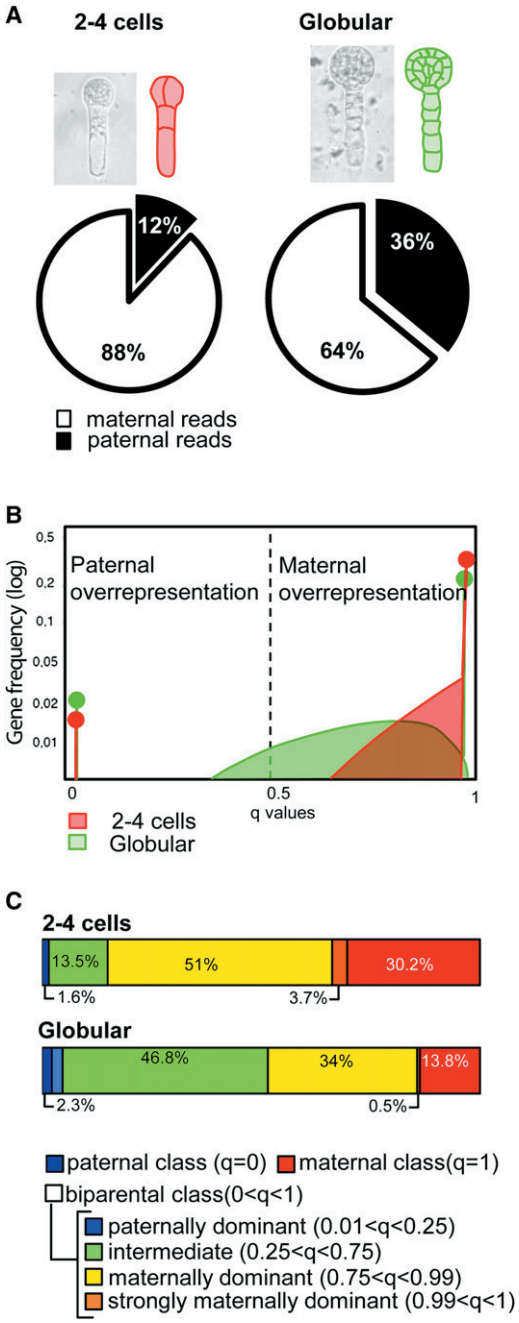


Figure 1. Parental Contributions to the *Arabidopsis* Early Embryo Transcriptome

(A) Distribution of maternal versus paternal reads covering the embryonic transcriptome of isolated embryos. Embryos were derived from a cross between polymorphic parents. Informative SOLiD reads (Table S1) mapping to known SNPs were assigned to the maternal (*Ler*) or paternal (Col) parent.

(B) Informative reads identified 3973 and 3078 genes in the 2–4 cell and globular transcriptomes, respectively. The graph shows a likelihood-based gene distribution according to the proportion of maternal transcripts, q. q = 1 and q = 0 represent genes contributed only maternally or paternally, respectively. Genes with 0 < q < 1 are contributed biparentally. y axis: proportion of genes (log scale) ; x axis:q values along 1% quantiles. The full-colored circles indicate extreme quantiles ($0 \leq q < 0.01$ and $0.99 < q \leq 1$).

(C) Composite diagram representation of the gene distribution as drawn in (B) according to q intervals as labeled. Related to Figure S1, Table S1, and Table S2.

2005; Vielle-Calzada et al., 2000). In contrast, a biparental expression in the zygote and early embryo was shown for certain other genes (Aw et al., 2010; Meyer and Scholten, 2007; Ronceret et al., 2005, 2008; Scholten et al., 2002; Weijers et al., 2001). Furthermore, near-saturation mutagenesis screens identified a plethora of mutations affecting embryo development at or before the globular stage. The majority of these segregate as zygotic recessive traits, indicating biparental contributions to early embryogenesis (Tzafrir et al., 2004). Thus, to date there is no clear understanding of the relative parental contributions to plant embryogenesis nor are the mechanisms regulating the respective contributions known.

We performed allele-specific profiling of the embryonic transcriptome and quantified relative transcript contributions of the paternal and maternal genomes in *Arabidopsis* embryos. We demonstrate a strong, genome-wide dominance of maternal transcripts at early stages, although many transcripts are biparentally represented. This finding reconciles the observation made by several laboratories of both maternal and zygotic effects during early embryogenesis. Using reporter and profiling analyses, we found an increasing contribution of paternal products as embryo development proceeds, with kinetics differing on a gene-by-gene basis. We identified two maternal epigenetic pathways, involving the chromatin siRNA pathway and the histone chaperone complex CAF1, which act antagonistically to regulate the paternal contribution. Importantly, we show that these pathways are distinct from those regulating genomic imprinting.

RESULTS

Analysis of Parental Contributions in Isolated *Arabidopsis* Embryos

To unambiguously define the parental contributions in the early embryo, we profiled the transcriptome of early-stage embryos in an allele-specific manner. We dissected 2–4 cell and globular stage embryos (Figure 1A) derived from a cross between the polymorphic Landsberg *erecta* (*Ler*) and Columbia (Col) accessions. The embryonic transcriptome was sequenced using a SOLiD v3 platform (Figure 1; see Table S1). Reads covering single-nucleotide polymorphisms (SNPs) (Borevitz et al., 2007) were extracted and are referred to hereafter as informative reads with respect to parental origin. Informative reads were used to quantify parental contributions both globally (Figure 1A), and for each gene individually. Genes were considered biparentally expressed at a given stage whenever both parental alleles were identified in the corresponding transcriptome. The classification of genes for which only a single allelic variant was detected was more ambiguous. Transcripts might be detected from only one parent because of uniparental expression or because of low sampling rates, which is of particular concern for genes expressed at low levels. To circumvent this difficulty, we made a probabilistic model describing the distribution of genes according to the proportion of maternal transcripts, q,

which was adjusted to best-fit the observations (Experimental Procedures). From this likelihood-based distribution, we estimated the frequency of genes fitting the biparental class ($0 < q < 1$, both parental alleles were detected), the paternal class ($q = 0$, only the paternal alleles were detected), or the maternal class ($q = 1$, only the maternal alleles were detected). In addition, each gene could be assigned a probability of falling within each class (Table S2). This integrated approach allowed a quantitative and qualitative analysis of parental contributions to the embryonic transcriptome.

The Transcriptome of 2–4 Cell Embryos Is Maternally Dominant despite Significant Contributions from Both Parental Genomes

Strikingly, at the 2–4 cell embryo stage 88.4% of the informative reads (n = 135,142) were of maternal origin (Figure 1A). This was confirmed in an independent biological replicate (Figure S1). The informative reads represented 3973 loci located throughout the genome. The gene distribution drawn for q quantiles showed a strong bias toward maternal overrepresentation (Figure 1B) with 85% of the genes described by $q > 0.75$ (Figure 1C; Table S1). Transcripts of the biparental class ($0 < q < 1$; 68.2% of the identified genes) contributed strongly to this maternal dominance with 54.7% genes described by $0.75 < q < 1$, i.e., maternal overrepresentation (Figure 1C and Table S1). Furthermore, our analysis revealed 30.2% transcripts of the maternal class ($q = 1$) against only 1.6% transcripts of the paternal class ($q = 0$) at the 2–4 cell stage. Thus, in *Arabidopsis* both parental genomes contribute to the early embryonic transcriptome but overall it is clearly dominated by maternal transcripts.

The Paternal Contribution Is Higher at the Globular Stage Concomitant with a Gradual Activation of Paternal Alleles

To determine how parental transcript contributions change during embryogenesis, we extended our allele-specific profiling to embryos at the globular stage. Although maternal dominance was maintained, the paternal contribution increased, as shown by 35.9% paternal reads versus 11.6% at the 2–4 cell stage (Figure 1A and Table S1). These informative reads identified 3078 loci, for which the q distribution remained skewed toward maternal overrepresentation, although to a lesser extent than at the 2–4 cell stage (Figure 1B). The maternal class ($q = 1$) represented 13.8% of the genes (versus 30.2% at the 2–4 cell stage), and the paternal class ($q = 0$) increased marginally to 2.3% (versus 1.6% at the 2–4 cell stage) (Table S1). Concomitantly, the biparental class increased, now comprising 83.9% of genes, with 34.5% showing maternal overrepresentation ($0.75 < q < 1$) (versus 54.7% at the 2–4 cell stage) (Figure 1C). Importantly, the globular stage transcriptome shared 2417 genes (78.5%) with that of the 2–4 cell stage, representing 95% of the reads (Figure 2A). We analyzed the changes of parental contributions among these shared genes by quantifying class transitions (Figure 2B). For instance a transition from the maternal class to the biparental or paternal class indicates de novo activation of the paternal allele and represents 21.5% (515) of the genes (Figure 2C). De novo activation of the maternal allele was less prominent with only 0.8% genes, mostly because few paternally expressed genes were identified at the 2–4 cell stage. This analysis identified loci with a decay of one parental transcript, with 2.5% and 9.7% showing loss of their maternal or paternal transcripts, respectively (Figure 2C). However, most quantitative changes occurred in the biparentally expressed class (54.4%, 1315 genes) where the majority showed a marked increase in the relative paternal contribution (778 genes, Figure 2D). This could result from decay of maternal RNAs, de novo transcription of paternal alleles, or a combination of both.

To refine the timing of activation of paternal alleles, we monitored paternal activity of six marker lines expressing a reporter gene under diverse promoters active during early embryogenesis. The lines reflect genes with diverse cellular functions (Table S3) and showed either early, intermediate, or late paternal activity, respectively (e.g., *RPS5A*, *CYCB1*;1, and *ET1041*), but did not appear in our allele-specific transcriptome due to the absence of a referenced SNP in their sequence. The marker lines clearly showed distinct expression depending on whether they were maternally or paternally inherited (Figure 3A and Figure S2A). We scored the number of F1 embryos showing paternal marker expression at the same developmental stage and in a wild-type maternal background (three to seven biological replicates each, Table S4). For all markers the proportion of stained embryos increased with developmental progression (Figure 3B and Figure S2B). Consistent with our RNA profiling results, paternal expression of the markers displayed gene-specific activation timing (developmental stage at which the first expression was detected) and kinetics (incremental increase in the fraction of progeny showing expression). For instance, in a L*er* maternal background the paternal *ET1041* marker showed only 4% stained embryos at the 2–4 cell stage, whereas the *RPS5A* and *GRP23* markers showed 58% and 67%, respectively (Figure 3B). In contrast, maternally transmitted markers showed consistent expression in essentially all embryos, even at earliest stages (Figures S2A and S2C).

Taken together with the transcriptome study, these findings strongly suggest that paternally inherited alleles—even those that are detectable at a very early stage—are activated gradually after fertilization. This is consistent with the previous observations made for individual genes in *Arabidopsis*, and reconciles earlier, apparently conflicting reports. Whether maternal loci follow similar or different activation kinetics cannot be easily resolved because of the potential importance of maternal carryover. Nevertheless, our observations are reminiscent of the gradual, de novo, expression of zygotic genes reported in animals (Tadros and Lipshitz, 2009).

Maternal KRYPTONITE Activity Controls Paternal Contribution in the Early Embryo

In our marker analyses we observed that expression was influenced by the maternal genotype (i.e., the accession) (Figure S2B). This indicated a maternal control of paternal expression as suggested previously (Ngo et al., 2007). The gradual increase in paternal allele expression might reflect the progressive release of a silencing mechanism. In *Arabidopsis*, silent chromatin is enriched in histone H3 dimethylated at lysine 9 (H3K9me2), a modification principally deposited by the SUVH4 histone methyltransferase KRYPTONITE (KYP) (Jackson et al.,

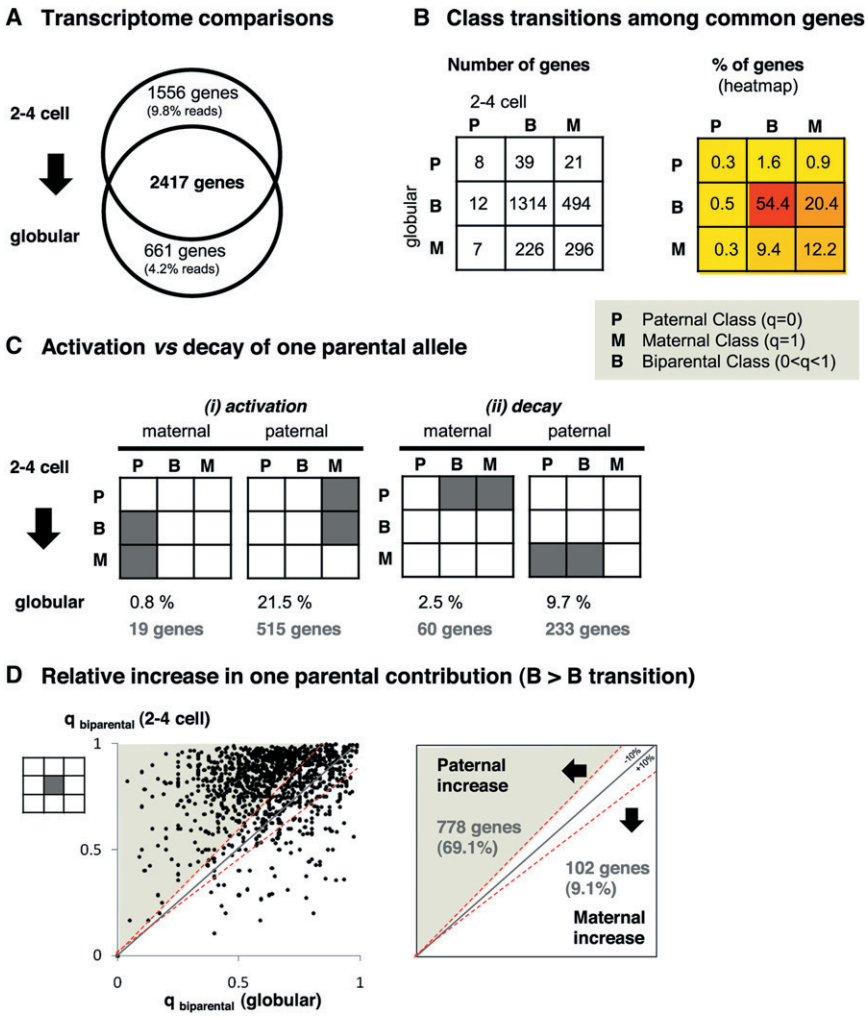


Figure 2. Dynamic Changes of Parental Contributions during Embryo Development

(A) Venn diagrams showing the number of genes identified by informative reads and shared between 2–4 cell and globular embryo transcriptomes. Global coverage for specific genes is indicated in brackets. Common genes are covered by the majority of reads (90%–95%).

(B) The transition tables describe the changes for common genes in the parental class (P, B, M, see legend) that occurred during development (2–4 cell → globular transition).

(C) Subsets of class transitions illustrate dynamic changes in allele representation as indicated (activation/de novo expression and decreased expression/decay of one parental allele) during developmental progression. Note that "decay" may correspond to a decrease in SNP coverage falling below detection threshold rather than an absolute loss of transcript. The % are the sum of the % genes in (B) falling into the gray transitions.

(D) A vast majority of common genes (54.4%) is biparentally represented at both stages. The proportion of maternal transcripts (q) was calculated for 1125 common genes sequenced on both alleles and plotted as indicated (left). The interpretation of relative changes toward higher paternal or maternal representation is shown (right). The inset shows the number of genes with changes in q values > 10% (dashed lines). A total of 245 genes showed no or <10% change. Related to Table S1 and Table S2.

2002). To investigate the possibility that maternal KYP regulates the activity of paternal alleles, we quantified reporter gene expression in seeds resulting from a cross between a maternal *kyp* mutant and wild-type pollen carrying the marker. For all reporters tested, lack of maternal KYP activity significantly increased the proportion of embryos showing early paternal reporter activity (before the 16-cell embryo stage) (Figures 4A and 4B, Figure S2A, and Table S4). We confirmed the maternal effect of the *kyp* mutation on the endogenous *AGP18* locus by allele-specific RT-PCR (Figure 4C). These data strongly suggest a role for maternal KYP activity in repressing the transcription of paternal alleles during early embryogenesis.

To investigate the maternal effect of the *kyp* mutation at the genome-wide level, we performed allele-specific profiling of 2–4 cell embryos dissected from crosses between maternal *kyp* (*Ler*) and paternal wild-type (Col) parents (Table S1). Embryos inheriting maternal *kyp* (kyp^m/KYP^p) showed a strong increase in the proportion of paternal reads (35.9% versus 11.6% in 2–4 cell stage wild-type embryos) (Figure 4D and replicate Figure S1) resulting in a similar paternal contribution as in wild-type embryos at the globular stage (36.1%; Figure 1A). Informative reads in kyp^m/KYP^p embryos identified 3125 genes (Table S1) for which the q distribution was shifted toward a higher paternal representation compared to wild-type embryos at the 2–4 cell stage (Figure 4E). Notably, the proportion of the paternal class increased markedly (10.3% at q = 0, versus 1.6% in wild-type 2–4 cell embryos, Figure 4F) whereas the proportion of genes in the maternal class diminished (Figure 4F). In addition, genes in the biparental class (0 < q < 1) showed a higher paternal contribution compared to wild-type embryos at the 2–4 cell stage (shifted distribution Figure 4E; less maternally dominant genes Figure 4F). Thus, the maternal *kyp* mutation affected a large number of genes throughout the genome, resulting in a higher paternal contribution to the early embryonic transcriptome.

Importantly, 2461 of these 3125 genes were common to the wild-type transcriptome of embryos at the same stage and were covered by 94.6% of the informative reads (Figure S3B). This suggests that the maternal *kyp* mutation does not drastically alter the 2–4 cell stage embryonic transcriptome, but instead modifies the relative parental contributions of genes normally expressed at this stage. Interestingly, the maternal *kyp* mutation induced class transitions similar to those induced by developmental progression (Figures S3C–S3E compared to Figures 2B–2D). The changes in the relative contribution of

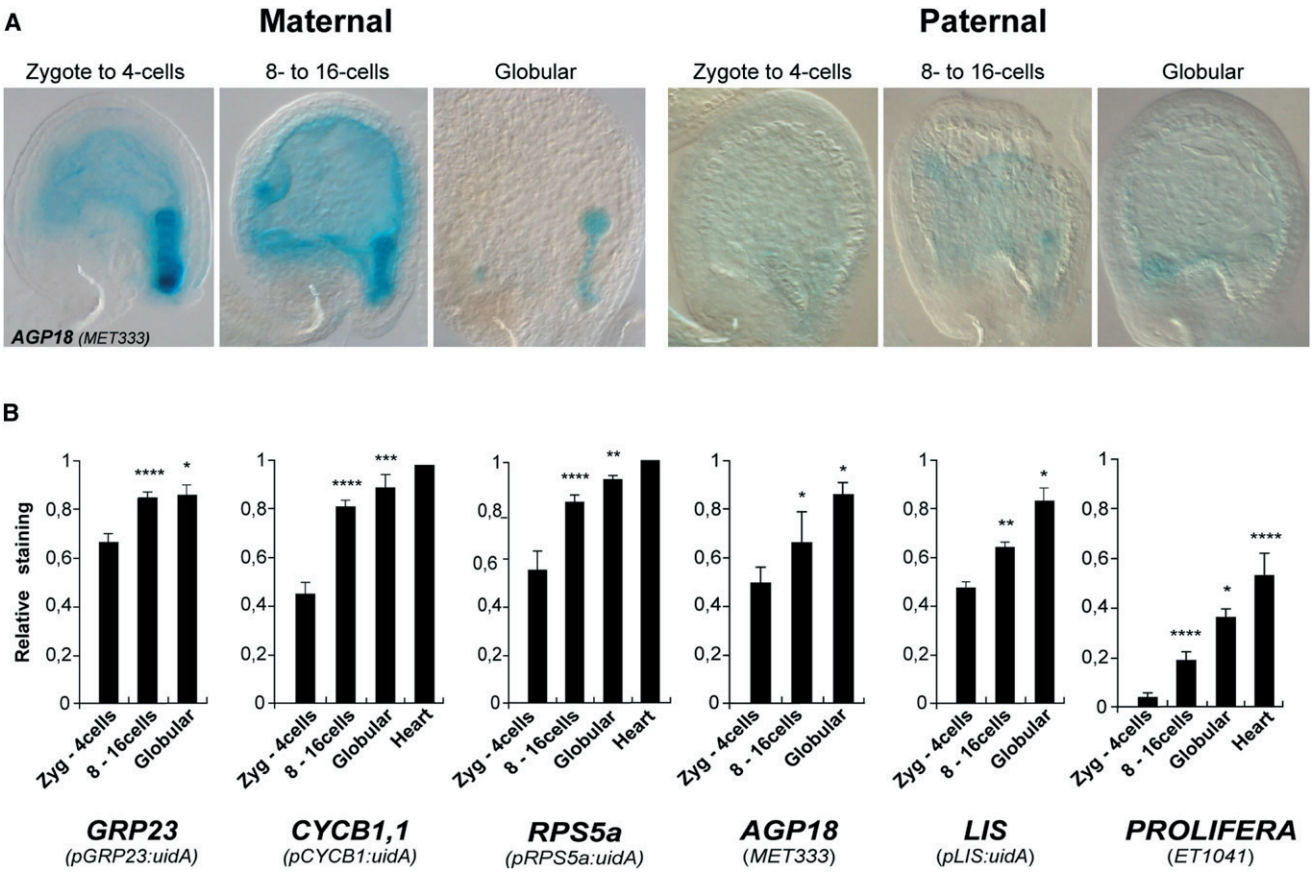


Figure 3. Gradual Activation of Paternal Markers during Early Embryo Development

(A) Representative panel showing differential expression of the marker tested (Table S3 and Table S4), here *MET333* reporting *AGP18* expression, when transmitted maternally (left) or paternally (right), as monitored by histochemical detection of GUS (blue substrate).

(B) Expression of the paternal markers was scored as the proportion of embryos showing GUS staining at a given developmental stage in a wild-type maternal background. Two-tailed Fisher's exact tests were used to assess differences between two consecutive developmental classes: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. Error bars represent standard error between independent biological replicates.

Related to Figure S2, Table S3, and Table S4.

biparental class genes were highly correlated (Figure 4G), except for a subgroup (158 genes) showing a higher increase of the paternal contribution in kyp^{m}/KYP^{p} 2–4 cell embryos as compared to wild-type globular embryos (Figure 4G). The analysis also identified genes showing a gain or a loss of one parental transcript (Figure S3D). We conservatively estimate that the maternal *kyp* mutation induces an increased paternal and maternal contribution for 1307 and 117 genes, respectively (Figure S3F).

Taken together these results strongly suggest that maternal KYP activity downregulates early transcription of many paternal, and some maternal, alleles of loci throughout the genome. This maternal control may be progressively released during embryogenesis, leading to the gradual activation of the paternal genome.

The Chromatin siRNA Pathway Maternally Controls the Paternal Contribution to the Early Embryo

Because KYP-dependent H3K9me2 in *Arabidopsis* is linked to non-CG DNA methylation, we verified the effects of mutations in *DOMAIN REARRANGED METHYLTRANSFERASE2* (*DRM2*) and *CHROMOMETHYLASE3* (*CMT3*), two genes that control non-CG methylation (Feng et al., 2010). Similar to *kyp*, both mutations inherited maternally resulted in precocious activation of the paternally inherited markers (Figure 5A and Figures S4A and S4B). By contrast, maternal mutations in either *METHYLTRANSFERASE1* (*MET1*) or *DECREASED DNA METHYLATION1* (*DDM1*), involved in the maintenance of CG DNA methylation (Feng et al., 2010), had no consistent effect (Figure S4C). Thus, non-CG but not CG DNA methylation participates in the transcriptional repression of paternal marker genes.

In *Arabidopsis*, DNA and histone methylation can be mediated by small-interfering RNAs (siRNAs) via the chromatin siRNA pathway, also known as the RNA-directed DNA methylation (RdDM) pathway (Brodersen and Voinnet, 2006). To determine whether the RdDM pathway plays a role in the maternal repression of paternal alleles, we tested mutations in the following RdDM components: *NRPD1a* (PolIV), *NRPD1b* (PolV/*NRPE1*), *DCL3*, *RDR2*, and *AGO4* (Table S3). When inherited maternally, all these mutations allowed an earlier and stronger detection of paternally transmitted markers (Figure 5A). By contrast, the *dcl2-1* mutant, which affects a distinct siRNA-dependent

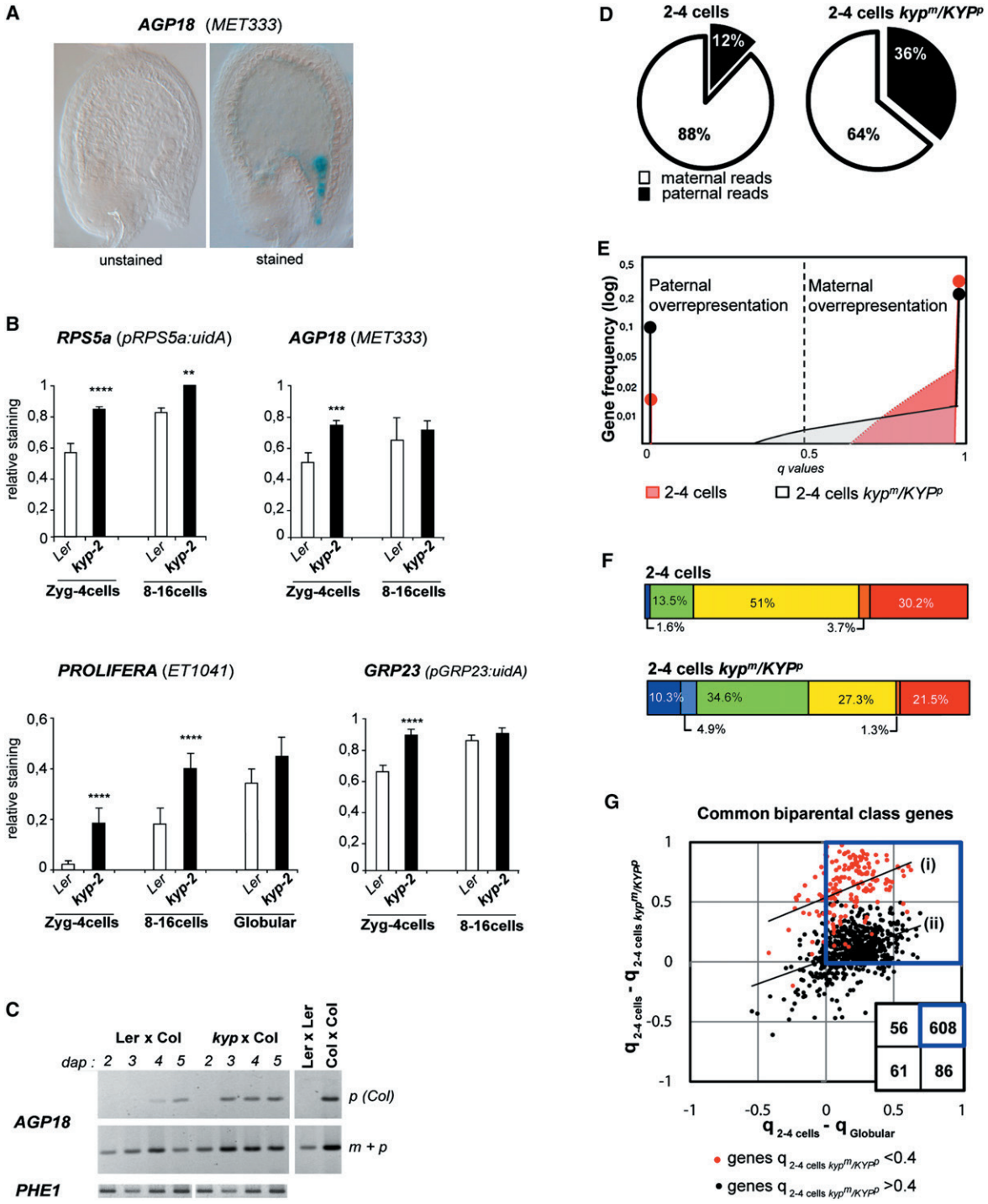


Figure 4. Maternal *KYP* Activity Controls Parental Contributions to the Early Embryo

(A) Histochemical staining for GUS activity from a paternally inherited embryo marker (*MET333*), showing the typical stained and unstained seeds as scored in the graphs in (B).

(B) Expression of paternal markers, in wild-type L*er* or *kyp-2* maternal backgrounds scored as described in Figure 3. See also Figure S3A, Table S3, and Table S4.

(C) Allele-specific RT-PCR of endogenous gene *AGP18* paternal transcripts in siliques harvested 1–5 days after pollination (dap) inheriting a maternal *kyp* mutation, as compared to the wild-type. Selective amplification of the paternal allele (p, top), amplification of maternal and paternal alleles (m + p, middle), control amplification of paternally expressed *PHE1* mRNA (bottom).

(D–G) Allele-specific transcriptome profiling in isolated 2–4 cells embryos inheriting a maternal *kyp-2* mutation compared to 2–4 cell wild-type embryos. Data and legends are as in Figure 1.

silencing pathway (Brodersen and Voinnet, 2006), did not alter the activation kinetics of the paternal reporters (Figure 5A). Importantly, paternal inheritance of mutant *KYP*, *CMT3*, or *NRPD1b* components showed no effect on paternal marker expression (Figure S4D), confirming a specific maternal role for the RdDM pathway in paternal marker regulation.

Several lines of evidence indicate that derepression of the paternal reporters was not linked to their transgenic nature. First, we confirmed precocious detection of paternal transcripts for the endogenous *AGP18* locus in embryos inheriting a maternal *kyp* mutation using allele-specific RT-PCR (Figure 4C). Second, several transgenic reporters under the control of transposon enhancers active in pollen (Slotkin et al., 2009) remained paternally undetectable in embryos inheriting a maternal *cmt3* mutation (Figure S4E), whereas ectopic maternal activation of the same reporters was reported in *cmt3* embryo sacs (Pillot et al., 2010a). Consistently, the profiling confirmed that the *kyp* mutation did not massively derepress transposons and repeats in the embryo (Table S1). Together with the genome-wide analysis of over 3000 endogenous loci in kyp^m/KYP^p embryos, these results indicate that the maternally inherited components of the RdDM pathway are involved in controlling the genome-wide transcriptional dynamics of paternally inherited alleles.

Consistent with the proposed global role for the RdDM pathway in transcriptional control in early embryos, we observed that *nrpd1a1b* mutant zygotes showed an abnormally high level of active PolII in their nucleus (Figure 5B). This was associated with abnormal deposition of the repressive H3K9me2 marks (Figure S5). The epigenetic and transcriptional states of zygotic nuclei in this RdDM mutant is thus in stark contrast to wild-type zygotes, which have a relatively quiescent transcriptional state (Pillot et al., 2010b).

Surprisingly, despite their effect on gene expression, RdDM mutants have not been reported to cause embryo lethality. However, this does not exclude subtle defects and, indeed, transient patterning defects were observed in embryos lacking maternal CMT3 activity (Pillot et al., 2010b). Similarly, early *kyp* embryos showed abnormal division planes in the embryo and suspensor cells, suggesting a role in early embryonic patterning (Figure S6).

Ovules Are Enriched in 24 nt siRNAs Targeting Gene-Coding Sequences

Our results suggest a novel role for the RdDM pathway in the regulation of genic regions, as this pathway had previously been associated mostly with transposon and repeat silencing. To verify the presence of maternal small RNAs targeting protein-coding regions, we profiled a library of small RNAs generated from manually dissected mature ovules before fertilization. We reasoned that if this were a maintenance mechanism, the siRNAs had to be produced before fertilization. Our analysis revealed a large fraction of siRNAs targeting genic regions (comprising protein-coding sequences [CDS] and 500 bp of putative 5′ regulatory regions of genes) in ovules as compared to whole inflorescence (Lu et al., 2005) (Figures 6A and 6B). This increase was not due to 21 nucleotide (nt) sRNAs (represented mainly by DCL1-dependent miRNAs and siRNAs) but was associated with the 24 nt fraction, whose biogenesis is dependent on PolIV, RDR2, and DCL3 (Brodersen and Voinnet, 2006). Overrepresentation of 24 nt siRNAs derived from CDS in ovules was correlated with the transcriptional control of individual loci mediated by maternal KYP activity in the embryo: genes showing a transition from maternal expression in the wild-type to biallelic or paternal expression in kyp^m/KYP^p embryos (*kyp*-responsive genes; Figure 6C) showed significantly more matching siRNAs than genes that remained maternally expressed in kyp^m/KYP^p embryos (*kyp*-unresponsive; Figure 6C). This finding is consistent with a role of the 24 nt siRNAs in regulating paternally inherited alleles.

Maternal CAF-1 and Histone H3 Turnover Regulate Transcriptional Activation of Paternal Alleles

Mutant analyses showed that lack of maternal RdDM components induced the precocious transcriptional activation of paternal alleles. It is unknown whether paternal alleles are in a state permissive to transcription or whether additional epigenetic reprogramming events are necessary for their activation. In the course of our genetic screen for mutants affecting paternal reporter expression, we identified MULTIPLE SUPPRESSOR OF IRA1 (MSI1) (Hennig et al., 2003) to be necessary for their expression. In embryos inheriting a maternal *msi1* mutation, the activation of paternal reporter genes was markedly delayed (Figures 7A and 7B and Figure S7A) compared to wild-type embryos. Reduction of paternal transcript levels in *msi1* mutants was confirmed by RT-PCR for the endogenous *GRP23* locus (Figure 7C).

MSI1 participates in several protein complexes including the CAF1 complex, a component of chromatin organization ensuring mitotic stability (Ono et al., 2006). CAF1 is formed by FASCIATA1 (FAS1), FAS2, and MSI1, and functions as an H3/H4-specific chaperone facilitating nucleosome assembly during replication (Hennig et al., 2005; Kaya et al., 2001). Consistently, maternal loss of another subunit of the CAF1 complex, FAS2, showed a similar effect as *msi1* (Figure 7A and Figure S7B). MSI1 is also a subunit of the MEA-FIE *Polycomb* group (PcG) complex active in seeds (Köhler et al., 2003), but a mutation affecting

(D) Distribution of maternal versus paternal reads as in Figure 1A.
(E) q distribution as in Figure 1B. A total of 3125 genes were identified by informative reads in mutant embryos.
(F) Composite diagram representation of the gene distribution in (E) as in Figure 1C.
(G) Scatter plot distribution of 811 biparental-class genes commonly detected in wild-type 2–4 cell embryos, globular embryos, and 2–4 cell kyp^m/KYP^p embryos. The difference in maternal contribution between stages ($q_{globular} - q_{2–4\ cell}$) or genotype ($q_{2–4\ cell\ WT} - q_{2–4\ cell\ kyp/KYP}$) is plotted on the x and y axis, respectively. q values were calculated for each transcript as follows: maternal reads/maternal + paternal reads. Differences >0 mean a higher paternal contribution compared to wild-type 2-4 cell embryos. Blue frame: genes with a correlated increased paternal contribution in both globular and kyp^m/KYP^p 2–4 cell embryos. Two groups of genes are delineated according to their relative maternal contribution (q) in kyp^m/KYP^p embryos as indicated (red, blue). Linear regressions: (i), $y = 0.47x + 0.54$; $R^2 = 0.17$; (158 genes); and (ii), $y = 0.41x + 0.02$; $R^2 = 0.20$ (653 genes).
Related to Figure S1, Figures S3B–S3F, Table S1, and Table S2.

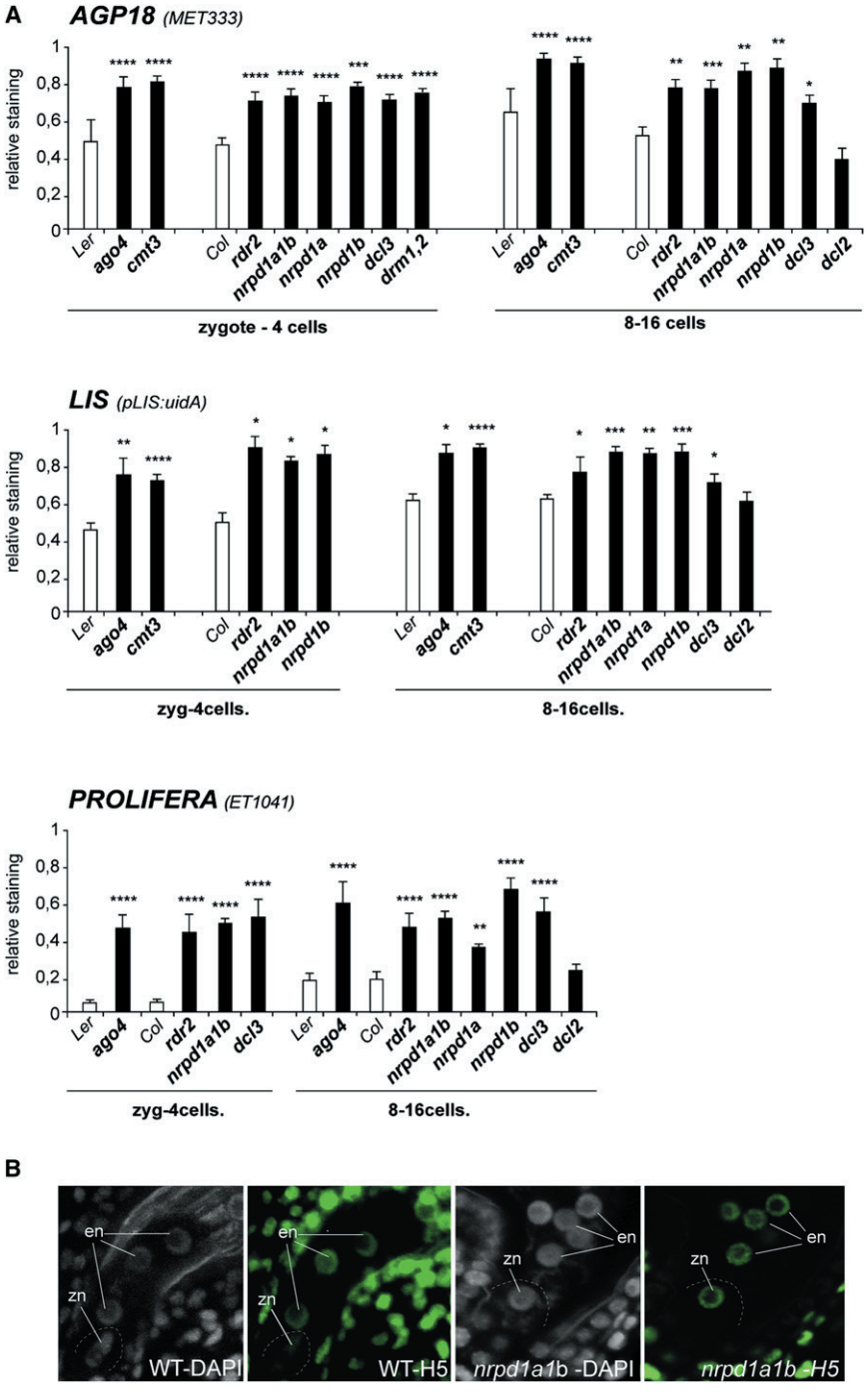


Figure 5. A Functional RdDM Pathway Is Required for Maternal Control of Paternal Embryo Markers and Maintenance of Transcriptional Quiescence in the Zygote

(A) Expression of paternal markers in embryos, scored as in Figure 3 and Figure 4, in maternal mutant backgrounds lacking activity of RdDM components as indicated. See also Figure S4, Figure S6, Table S3, Table S4, and Table S5.

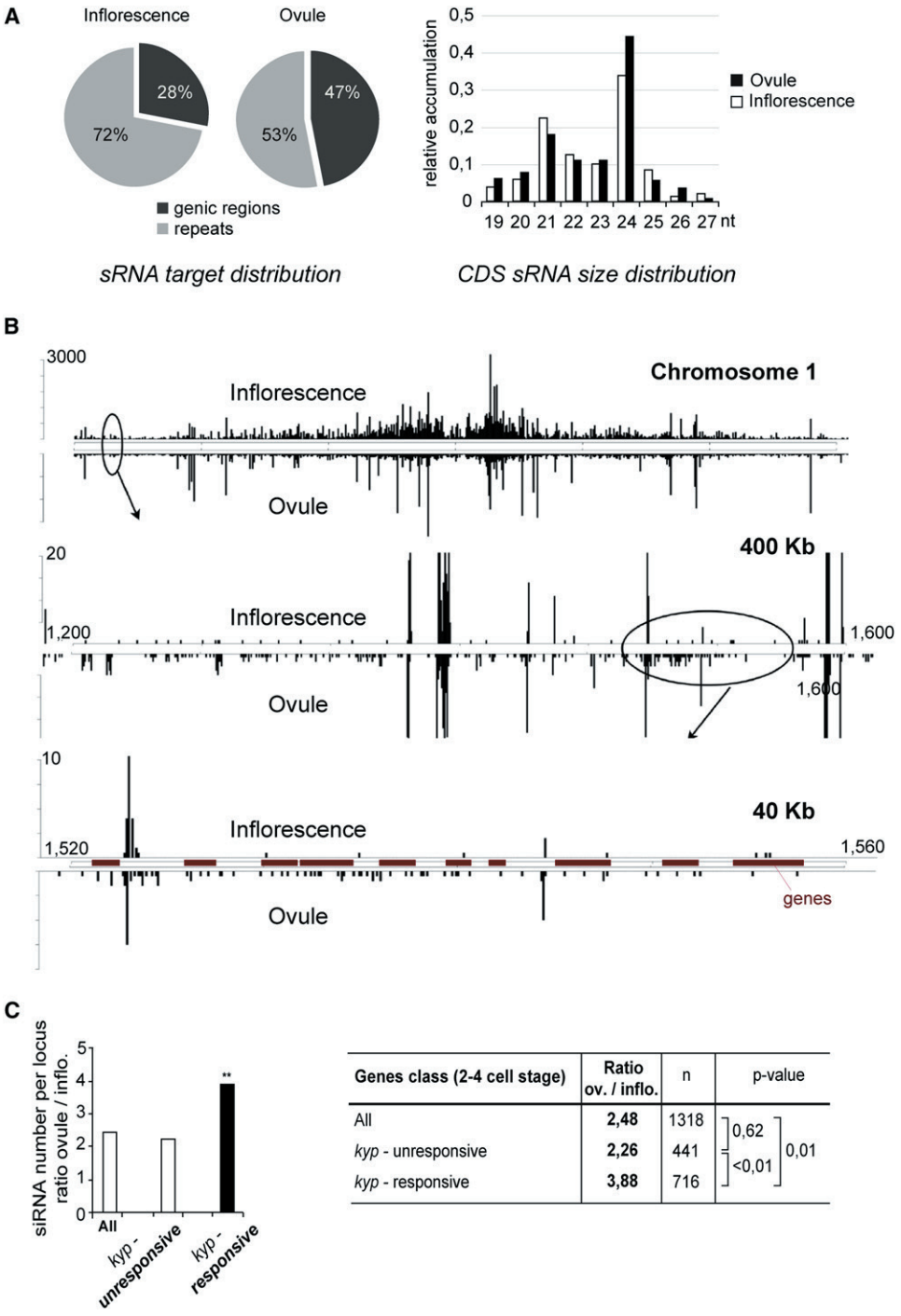
(B) Immunolocalization of the active form of RNA polymerase II (H5) in PolIV/PolV mutant zygotes (*nrpd1a1b*) compared to the wild-type (WT). DNA was counterstained with DAPI. See also Figure S5.

the MEDEA (MEA) subunit had no effect on paternal expression at the globular stage (Figure S7C). These results strongly suggest that the CAF1 complex is maternally required to activate transcription of the paternal genome, likely via histone turnover. CAF1 may regulate the incorporation of specific histone variants controlling transcriptional activity in plants, as it is the case in animals. For example, in the animal germ line H3.3 variants are incorporated at actively transcribed loci (Ooi et al., 2006).

A similar role may be proposed in plants, because we observed that maternal mutations affecting the H3.3 variants HTR4 and HTR5 (Okada et al., 2005) significantly delayed the activation of paternal markers in the embryo (Figure S7D). These results indicate a role for CAF1 and H3/H3.3 turnover in the transcriptional activation of paternal—and possibly maternal—alleles after fertilization.

DISCUSSION

We provided a genome-wide view of the parental contributions to early embryogenesis in *Arabidopsis*. At early stages, the maternal transcriptome clearly predominates: although 68% of the genes are biparentally expressed, their maternal transcripts are overrepresented. Further, >30% of the genes show an exclusively maternal contribution. During the transition from the 2–4 cell to the globular stage the paternal contribution increases. We showed that these parental contributions are maternally controlled by two antagonistic regulatory pathways regulating the onset of paternal and, at least partially, maternal zygotic transcription. Maternal dominance at early stages results from downregulation of paternal alleles at loci throughout the genome via the chromatin siRNA pathway, linked to RNA-directed DNA and histone methylation. In addition, transcriptional activation of paternal alleles involves histone exchange, possibly via the replacement of H3.3 variants, for which rapid turnover is observed after fertilization (Ingouff et al., 2010). Release of silencing might also involve passive DNA-demethylation during mitoses or the activity of DNA- or histone-demethylases. Although additional investigations are required to refine the mechanistic role of these events in the control of the zygotic genome, our results suggest that flowering plants evolved



Genes class (2-4 cell stage)	Ratio ov. / inflo.	n	p-value
All	**2,48**	1318	0,62
kyp - unresponsive	**2,26**	441	<0,01
kyp - responsive	**3,88**	716	0,01

Figure 6. Ovules Are Enriched in Small RNAs Targeting Genic Regions

(A) Deep sequencing of small RNA (sRNAs) from mature ovules reveals increased targeting to genic regions (comprising protein coding sequences (CDS) and putative 5′-regulatory regions 500 bp upstream of ATG) as compared to sRNAs from inflorescence (pie charts). Size distribution of CDS-targeted sRNAs showed an increase in the 24 nt fraction in ovules as compared to inflorescence sRNA libraries (histogram).

(B) Comparative mapping of sRNA distribution in inflorescence (upper line) and ovules (lower line) libraries, exemplified by chromosome 1 (top), with a 400 kb zoom (middle), and 40 kb zoom (bottom). Boxes represent protein coding units (genes). CDS-specific 24 nt sRNA were distributed between + and − strands in the proportion of 63(+):37(−) in the ovule library.

(C) Average number of 24 nt siRNA per locus for maternal- and biparental-class genes showing no significant changes (*kyp*-unresponsive genes) or reduced q values (increased paternal contribution) in kyp^{m}/KYP^{p} embryos compared to wild-type (*kyp*-responsive genes). t test p values (table and graph, *p < 0.01) refers to differences in siRNA mean number between groups.

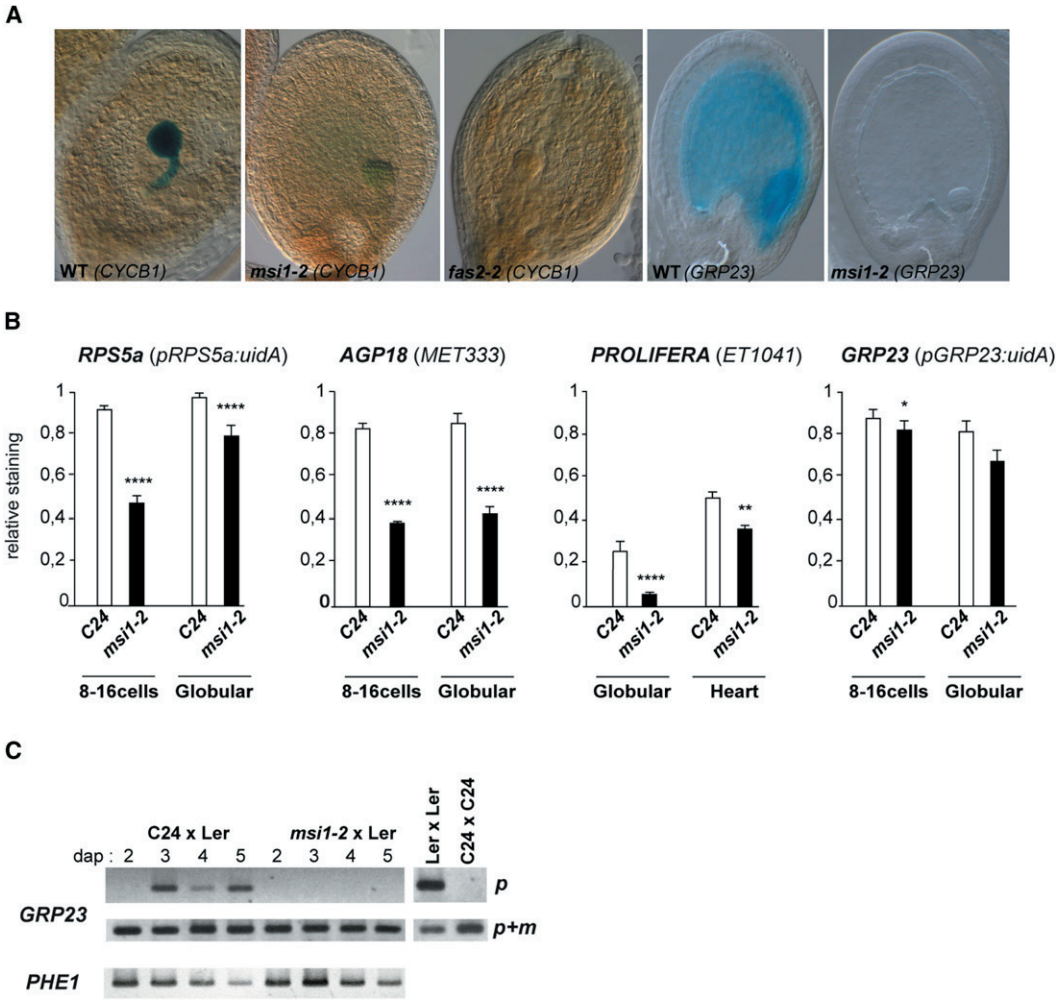


Figure 7. The Maternal CAF1 Nucleosome Assembly Complex Controls Paternal Gene Activation

(A) Paternal marker expression in *msi1* or *fas2* maternally mutant embryos compared to wild-type.

(B) Paternal marker expression scored as in Figure 3, Figure 4, and Figure 5, in *msi1* or *fas2* maternal background. See also Figure S7, Table S3, Table S4, and Table S5.

(C) Allele-specific RT-PCR shows reduced endogenous *GRP23* paternal transcript levels in siliques collected 2–5 dap inheriting a maternal *msi1-2* mutation, as compared to the wild-type. Selective amplification of the paternal allele (p), or of both maternal and paternal alleles (m + p), control amplification of paternally expressed *PHE1*.

strategies to regulate early embryogenesis that are similar to those described in animals (Baroux et al., 2008; Tadros and Lipshitz, 2009).

Maternal Effect and Zygotic Functions during Early Embryogenesis

The relative contribution and dynamic changes of parental transcripts we observed in early *Arabidopsis* embryos may result from a combination of de novo transcription postfertilization and transcripts carried over from the egg or sperm, although their respective abundance is unknown. 1331 and 621 genes with respective maternal and paternal contributions (of 3399 and 1482 represented on the ATH1 microarray) were consistently detected as present in egg- or sperm-specific microarray experiments (Borges et al., 2008; Wuest et al., 2010). Although these overlaps between pre- and postfertilization transcriptomes indicate a carryover from egg and sperm cells, potentially influencing early embryo development (Bayer et al., 2009; Pillot et al., 2010b), the expression status of these genes in the early embryo remains unknown. For many loci, stored and de novo expressed transcripts likely coexist during early embryogenesis (Tadros and Lipshitz, 2009), as it was shown by chromosomal deletions in *Drosophila* (De Renzis et al., 2007). Whenever exactly these genes may be expressed, the vast majority of transcripts are derived from the maternal genome, providing extensive maternal control over early development, explaining the existence of numerous maternal effect genes (reviewed in Baroux et al., 2008). However, it is equally clear that there is a paternal contribution to early embryogenesis from many loci, sometimes exclusively, consistent with paternal effect and zygotic genes acting early after fertilization (Bayer

et al., 2009, Meyer and Scholten, 2007; Ronceret et al., 2005, 2008; Scholten et al., 2002; Tzafrir et al., 2004; Weijers et al., 2001).

Epigenetic Pathways Controlling Parental Contributions Are Distinct from Those Regulating Genomic Imprinting

In plants and mammals, certain genes are regulated by genomic imprinting and are expressed monoallelically only from one parental allele (reviewed in Raissig et al., 2011). One could imagine that for maternally expressed, imprinted loci, the RdDM pathway repressing paternal alleles stays in place throughout development, thus leading to monoallelic maternal expression. However, the maternal regulatory pathways we uncovered complement—and act beyond—the regulation of genes by genomic imprinting: neither *KYP* nor *CMT3* regulates the imprinted *FIS2* locus (Jullien et al., 2006b). Conversely, we determined that *dcl3*, a mutation affecting the RdDM pathway, does not alter paternal silencing of the imprinted *MEA* gene, nor do mutations affecting the CAF1 subunit FAS2 (Table S5). Instead, the MEA-FIE PcG complex maintains silencing of the paternal *MEA* allele via H3K27 methylation (Baroux et al., 2006; Jullien et al., 2006a). This complex is not required for global repression of the paternal genome because a mutation affecting the MSI1 PcG subunit did not show precocious activation of paternal markers, but instead had a delaying effect. Thus, the maternal mechanisms controlling the timing of paternal genome activation described here are distinct from those regulating genomic imprinting. Furthermore, although parental contributions were described in the late endosperm (Hsieh et al., 2011), their regulation at early stages still needs to be elucidated at a genome-wide level. Possibly, different transcriptional controls are in place because the endosperm is actively transcribed soon after fertilization (Aw et al., 2010; Pillot et al., 2010b), owing to its differential targeting by a global DNA demethylase pathway that may counteracts the RdDM pathway (Gehring et al., 2009; Hsieh et al., 2009).

The RdDM Pathway Plays a Role in Early Seed Development

We have shown that maternal mutations affecting the PolIV subunit NRPD1a de-repressed the activity of paternal markers and modified the transcriptional status of the zygotic genome. Thus, we propose that PolIV-dependent 24 nt maternal siRNAs epigenetically control the transcriptional status of paternal, and possibly also maternal, loci throughout the genome. Whether an epigenetic dimorphism establishes differential susceptibility of the parental alleles to siRNA-based regulation is a challenging question that remains to be addressed. Maternal siRNAs were recently detected at later stages of seed development in the endosperm (Mosher et al., 2009), and thus our findings extend their role to early stages of embryogenesis, particularly in regulating protein-coding sequences. The endosperm and its progenitor, the central cell, or alternatively maternal sporophytic tissue (Olmedo-Monfil et al., 2010) have recently been proposed as a potential source of mobile maternal siRNAs driving silencing in the egg cell and the embryo (reviewed, e.g., in Bourc'his and Voinnet, 2010; Feng et al., 2010), an attractive hypothesis awaiting confirmation.

Perturbation of the maternal RdDM pathway leads to transient patterning defects as reported for embryos lacking maternal CMT3 (Pillot et al., 2010b) or, as described here, KYP activity. Thus, the pathway seems to fine-tune the expression of early patterning genes. Alternatively, phenotypic aberrations might be revealed in embryos inheriting a divergent paternal genome distinct from the maternal background, which provides the epigenetic control on zygotic genome expression. Further studies are awaited to determine if RdDM mutants might represent sensitized backgrounds to hybridization and how these maternal pathways affect out-breeding species. Maternal siRNAs, particularly those targeting transposable elements, have been proposed to act in heterosis and inter-specific hybridization (Chen, 2010; Martienssen, 2010). Twenty-four nucleotide siRNAs targeting coding regions are also downregulated in hybrid offspring (Groszmann et al., 2011). Our data suggest that siRNA-based mechanisms also target protein-coding sequences of the early embryonic genome, to control chromatin-based parental interactions during the epigenetic reprogramming that occurs after fertilization.

EXPERIMENTAL PROCEDURES

Plant Material

Arabidopsis thaliana accessions Columbia-0 (Col), Landsberg *erecta* (L*er*), C24, WS or Nossen (No) were used as wild-type controls according to the mutant's background. The marker lines and mutants are listed in Table S3 and genotyping assays are described in Extended Experimental Procedures.

Embryonic cDNA Libraries, Sequencing, and Allele-Specific Transcriptome Analysis

The full method is described in Extended Experimental Procedures. In brief, embryos were released from the seeds by gentle pressure and isolated under an inverted microscope using a microcapillary. Total RNA was extracted using the PicoPure RNA Isolation Kit (Arcturus) and 300–700 pg was amplified in a linear fashion using the WT-Ovation Pico RNA Amplification System (NuGEN Technologies). After second strand cDNA synthesis and library preparation, 50 bases sequence reads were generated by SOLiD v3 (Applied Biosystems) and aligned to the *Arabidopsis* Col genome (TAIR8.0). Unique reads mapping full length in the exome were sorted for the presence of L*er* annotated polymorphisms (Borevitz et al., 2007).

The analysis of parental contributions was performed using a likelihood-based model fitting best the observed distribution of maternal and paternal reads per transcripts. A detailed explanation is provided in Extended Experimental Procedures.

Marker Gene Analysis

The activity of paternal markers following crosses to wild-type or mutant females was assayed by histochemical staining of the *uidA* reporter gene product (the GUS enzyme) as described in Extended Experimental Procedures. The number of GUS-positive progeny was scored for each developmental stage and differences were assessed using two-tailed Fisher's exact test. For allele-specific RT-PCR, LNA-modified primers targeting a SNP from one parental transcript were used. Details on the reactions and primer sequences are provided in Extended Experimental Procedures.

Profiling of Small RNAs in Ovules

In brief, total RNA was extracted from dissected mature ovules (Peiffer et al., 2008) and a small RNA library was prepared and sequenced using Illumina Genome Analyzer. After filtering, reads were mapped against *Arabidopsis* Col reference sequence (TAIR 8) and compared to inflorescence small RNA reads (Lu et al., 2005), analyzed using the same procedure. Mapping, occurrence information, normalization, and graphical displays were computed using R. Genic regions and repeat target analysis was done using TAIR 8 genome release and ftp://ftpmips.helmholtz-muenchen.de/plants/cress/.

Whole-Mount Immunolocalization

Immunodetection was performed essentially as described (Pillot et al., 2010b), using antibodies from Abcam: anti-H3K9me2 (#ab1220) and anti-phosphoS2 RNA PolII [H5] (#ab24758) specific to the active form of PolII (Palancade and Bensaude, 2003). Images were captured on a laser scanning confocal microscope (Leica SP2) and maximum-intensity projections of selected optical sections generated. Details are provided in Extended Experimental Procedures.

ACCESSION NUMBERS

Data from embryo SOLiD profiling are accessible under GEO accession number GSE24198. Data from ovule small RNA profiling are accessible under GEO accession number GSE28627. Sequence data from this article can be found in the *Arabidopsis* Genome Initiative or GenBank/EMBL databases under the following accession numbers: At5g13960 (*KRYPTONITE*), At5g14620 (*DRM2*), At5g15380 (*DRM1*), At1g69770 (*CMT3*), At5g49160 (*MET1*), At5g66750 (*DDM1*), At4g11130 (*RDR2*), At3g43920 (*DCL3*), At3g03300 (*DCL2*), At2g27040 (*AGO4*), At1g63020 (*NRPD1A/POLIVA*), At2g40030 (*NRPD1b/NRPE1/POLV*), At5g58230 (*MSI1*), At5g64630 (*FAS1*), At5g64630 (*FAS2*), At1g02580 (*MEDEA*), At4g40030 (*HTR4*), At4g40040 (*HTR5*), At4g37450 (*AGP18*), At4g02060 (*PROLIFERA*), At4g37490 (*CYCB1;1*), At3g11940 (*RPS5a*), At2g241500 (*LACHESIS*), At1g10270 (*GRP23*).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, five tables, and one movie and can be found with this article online at doi:10.1016/j.cell.2011.04.014.

ACKNOWLEDGMENTS

We thank the *Arabidopsis* Stock Centers (NASC/ABRC), J. Carrington, R. Gross-Hardt, R. Martienssen, F. Meins, T. Lagrange, J. Paszowski, and E. Richards for seeds; C. Michaud, N. Duran-Figueroa, and V. Olmedo-Monfil for technical support; T. Nawy for advice on embryo microdissection; S. Schreiber, L. Bossen, and M. Schilhabel for sequencing support at Kiel; and M. Schmid for help in GEO data submission. We are grateful to M. Arteaga-Vazquez, O. Hamant, and T. Lagrange for helpful comments on earlier versions of the manuscript and to S. Kessler and S. Gillmor for critical reading. C.B., M.R., S.G., and A.S. were supported by the University of Zürich and grants of Swiss National Science Foundation and the European Union (FP5 ApoTool Project) to U.G. D.A., O.L., and D.G. were supported by the Institut de Recherche pour le Développement and the Agence Nationale de la Recherche, and T.L. by CNRS and a starting grant of the European Research Council. P.R. is supported by the Clusters of Excellence: "The Future Ocean" and "Inflammation at Interfaces." J.-P.V.-C. is supported by CONACyT and is an International Scholar of the Howard Hughes Medical Institute.

Received: January 21, 2011
Revised: March 28, 2011
Accepted: April 15, 2011
Published: May 26, 2011

REFERENCES

Andéol, Y. (1994). Early transcription in different animal species: implication for transition from maternal to zygotic control in development. Roux. Arch. Dev. Biol. *204*, 3–10.

Aw, S.J., Hamamura, Y., Chen, Z., Schnittger, A., and Berger, F. (2010). Sperm entry is sufficient to trigger division of the central cell but the paternal genome is required for endosperm development in Arabidopsis. Development *137*, 2683–2690.

Baroux, C., Autran, D., Gillmor, C.S., Grimanelli, D., and Grossniklaus, U. (2008). The maternal to zygotic transition in animals and plants. Cold Spring Harb. Symp. Quant. Biol. *73*, 89–100.

Baroux, C., Blanvillain, R., and Gallois, P. (2001). Paternally inherited transgenes are down-regulated but retain low activity during early embryogenesis in Arabidopsis. FEBS Lett. *509*, 11–16.

Baroux, C., Gagliardini, V., Page, D.R., and Grossniklaus, U. (2006). Dynamic regulatory interactions of Polycomb group genes: MEDEA autoregulation is required for imprinted gene expression in Arabidopsis. Genes Dev. *20*, 1081–1086.

Bayer, M., Nawy, T., Giglione, C., Galli, M., Meinnel, T., and Lukowitz, W. (2009). Paternal control of embryonic patterning in Arabidopsis thaliana. Science *323*, 1485–1488.

Borevitz, J.O., Hazen, S.P., Michael, T.P., Morris, G.P., Baxter, I.R., Hu, T.T., Chen, H., Werner, J.D., Nordborg, M., Salt, D.E., et al. (2007). Genome-wide patterns of single-feature polymorphisms in Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA *104*, 12057–12062.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijo, J.A., and Becker, J.D. (2008). Comparative transcriptomics of Arabidopsis sperm cells. Plant Physiol. *148*, 1168–1181.

Bourc'his, D., and Voinnet, O. (2010). A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. Science *330*, 617–622.

Brodersen, P., and Voinnet, O. (2006). The diversity of RNA silencing pathways in plants. Trends Genet. *22*, 268–280.

Bushell, C., Spielman, M., and Scott, R.J. (2003). The basis of natural and artificial postzygotic hybridization barriers in Arabidopsis species. Plant Cell *15*, 1430–1442.

Chaudhury, A.M., and Berger, F. (2001). Maternal control of seed development. Semin. Cell Dev. Biol. *12*, 381–386.

Chen, Z.J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci. *15*, 57–71.

De Renzis, S., Elemento, O., Tavazoie, S., and Wieschaus, E.F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. PLoS Biol. *5*, e117.

Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science *330*, 622–627.

Gehring, M., Bubb, K.L., and Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science *324*, 1447–1451.

Grimanelli, D., Perotti, E., Ramirez, J., and Leblanc, O. (2005). Timing of the maternal-to-zygotic transition during early seed development in maize. Plant Cell *17*, 1061–1072.

Groszmann, M., Greaves, I.K., Albertyn, Z.I., Scofield, G.N., Peacock, W.J., and Dennis, E.S. (2011). Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. Proc. Natl. Acad. Sci. USA *108*, 2617–2622.

Hennig, L., Bouveret, R., and Gruissem, W. (2005). MSI1-like proteins: an escort service for chromatin assembly and remodeling complexes. Trends Cell Biol. *15*, 295–302.

Hennig, L., Taranto, P., Walser, M., Schonrock, N., and Gruissem, W. (2003). Arabidopsis MSI1 is required for epigenetic maintenance of reproductive development. Development *130*, 2555–2565.

Hsieh, T.F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm. Science *324*, 1451–1454.

Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., et al. (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. Proc. Natl. Acad. Sci. USA *108*, 1755–1762.

Ingouff, M., Rademacher, S., Holec, S., Soljic, L., Xin, N., Readshaw, A., Foo, S.H., Lahouze, B., Sprunck, S., and Berger, F. (2010). Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in *Arabidopsis*. Curr. Biol. *20*, 2137–2143.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556–560.

Jahnke, S., Sarholz, B., Thiemann, A., Kuhr, V., Gutierrez-Marcos, J.F., Geiger, H.H., Piepho, H.P., and Scholten, S. (2010). Heterosis in early seed development: a comparative study of F1 embryo and endosperm tissues 6 days after fertilization. Theor. Appl. Genet. *120*, 389–400.

Jullien, P.E., Katz, A., Oliva, M., Ohad, N., and Berger, F. (2006a). Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in *Arabidopsis*. Curr. Biol. *16*, 486–492.

Jullien, P.E., Kinoshita, T., Ohad, N., and Berger, F. (2006b). Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. Plant Cell *18*, 1360–1372.

Kaya, H., Shibahara, K.I., Taoka, K.I., Iwabuchi, M., Stillman, B., and Araki, T. (2001). FASCIATA genes for Chromatin Assembly Factor-1 in Arabidopsis maintain the cellular organization of apical meristems. Cell *104*, 131–142.

Köhler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., and Gruissem, W. (2003). Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. EMBO J. *22*, 4804–4814.

Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C., and Green, P.J. (2005). Elucidation of the small RNA component of the transcriptome. Science *309*, 1567–1569.

Martienssen, R.A. (2010). Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of Arabidopsis. New Phytol. *186*, 46–53.

Meyer, S., and Scholten, S. (2007). Equivalent parental contribution to early plant zygotic development. Curr. Biol. *17*, 1686–1691.

Mosher, R.A., Melnyk, C.W., Kelly, K.A., Dunn, R.M., Studholme, D.J., and Baulcombe, D.C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature *460*, 283–286.

Ngo, Q.A., Moore, J.M., Baskar, R., Grossniklaus, U., and Sundaresan, V. (2007). Arabidopsis GLAUCE promotes fertilization-independent endosperm development and expression of paternally inherited alleles. Development *134*, 4107–4117.

Okada, T., Endo, M., Singh, M.B., and Bhalla, P.L. (2005). Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. Plant J. *44*, 557–568.

Olmedo-Monfil, V., Duran-Figueroa, N., Arteaga-Vazquez, M., Demesa-Arevalo, E., Autran, D., Grimanelli, D., Slotkin, R.K., Martienssen, R.A., and Vielle-Calzada, J.P. (2010). Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature *464*, 628–632.

Ono, T., Kaya, H., Takeda, S., Abe, M., Ogawa, Y., Kato, M., Kakutani, T., Mittelsten Scheid, O., Araki, T., and Shibahara, K. (2006). Chromatin Assembly Factor 1 ensures the stable maintenance of silent chromatin states in Arabidopsis. Genes Cells *11*, 153–162.

Ooi, S.L., Priess, J.R., and Henikoff, S. (2006). Histone H3.3 variant dynamics in the germline of Caenorhabditis elegans. PLoS Genet. *2*, e97.

Pahlavani, M.H., and Abolhasani, K. (2006). Xenia effect on seed and embryo size in cotton (Gossypium hirsutum L.). J. Appl. Genet. *47*, 331–335.

Palancade, B., and Bensaude, O. (2003). Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. Eur. J. Biochem. *270*, 3859–3870.

Peiffer, J.A., Kaushik, S., Sakai, H., Arteaga-Vazquez, M., Sanchez-Leon, N., Ghazal, H., Vielle-Calzada, J.P., and Meyers, B.C. (2008). A spatial dissection of the Arabidopsis floral transcriptome by MPSS. BMC Plant Biol. *8*, 43.

Pillot, M., Autran, D., Leblanc, O., and Grimanelli, D. (2010a). A role for CHROMOMETHYLASE3 in mediating transposon and euchromatin silencing during egg cell reprogramming in Arabidopsis. Plant Signal. Behav. *5*, 1167–1170.

Pillot, M., Baroux, C., Vazquez, M.A., Autran, D., Leblanc, O., Vielle-Calzada, J.P., Grossniklaus, U., and Grimanelli, D. (2010b). Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in Arabidopsis. Plant Cell *22*, 307–320.

Raissig, M.T., Baroux, C., and Grossniklaus, U. (2011). Regulation and flexibility of genomic imprinting during seed development. Plant Cell *23*, 16–26.

Ronceret, A., Guilleminot, J., Lincker, F., Gadea-Vacas, J., Delorme, V., Bechtold, N., Pelletier, G., Delseny, M., Chaboute, M.E., and Devic, M. (2005). Genetic analysis of two Arabidopsis DNA polymerase epsilon subunits during early embryogenesis. Plant J. *44*, 223–236.

Ronceret, A., Gadea-Vacas, J., Guilleminot, J., Lincker, F., Delorme, V., Lahmy, S., Pelletier, G., Chaboute, M.E., and Devic, M. (2008). The first zygotic division in Arabidopsis requires de novo transcription of thymidylate kinase. Plant J. *53*, 776–789.

Scholten, S., Lörz, H., and Kranz, E. (2002). Paternal mRNA and protein synthesis coincides with male chromatin decondensation in maize zygotes. Plant J. *32*, 221–231.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell *136*, 461–472.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Development *136*, 3033–3042.

Tzafrir, I., Pena-Muralla, R., Dickerman, A., Berg, M., Rogers, R., Hutchens, S., Sweeney, T.C., McElver, J., Aux, G., Patton, D., et al. (2004). Identification of genes required for embryo development in Arabidopsis. Plant Physiol. *135*, 1206–1220.

Vielle-Calzada, J.P., Baskar, R., and Grossniklaus, U. (2000). Delayed activation of the paternal genome during seed development. Nature *404*, 91–94.

Weijers, D., Geldner, N., Offringa, R., and Jürgens, G. (2001). Seed development: early paternal gene activity in Arabidopsis. Nature *414*, 709–710.

Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenfuhrer, J., von Mering, C., and Grossniklaus, U. (2010). Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. Curr. Biol. *20*, 506–512.

CHAPTER 2.2.2. *Arabidopsis* Embryogenesis - Maternal Dominance or Equal Parental Policy?

We have previously shown that the transcriptome of *Arabidopsis* embryos derived from crosses between the accessions Landsberg *erecta* (Ler) and Columbia (Col-0) is largely dominated by maternal reads (88%) at early stages (2-4 cells). Despite this maternal dominance, 66% of the genes have transcripts from both parental alleles, consistent with the fact that many embryo lethal mutations with preglobular developmental phenotypes are zygotically recessive (Autran et al., 2011). Transcriptome analyses at the globular stage, in conjunction with expression analyses of seven reporter loci, confirmed a gradual increase of paternal transcripts during embryogenesis, reflecting progressive zygotic genome activation (ZGA; Autran et al., 2011). We also demonstrated that paternal loci are epigenetically regulated by two antagonistic maternal pathways: a siRNA-based mechanism involving genes of the RNA-dependent DNA methylation (RdDM) pathway restricts expression of paternal alleles, while their activation relies on a nucleosome-remodeling pathway (Autran et al., 2011). As a result, *kyp*^{m}/*KYP*^{p} maternally mutant embryos show both a higher proportion of paternal reads (34% *vs* 12% in the wild type) and a gene distribution, based on a statistical best-fit model, that is skewed towards higher paternal contributions (Autran et al., 2011).

In contrast, a recent study using *Arabidopsis* embryos derived from crosses between the accessions Cape Verde Island (Cvi) and Col-0, showed a transcriptome with an equal contribution of paternal and maternal transcripts (Nodine and Bartel, 2012). To explain this discrepancy, the authors suggested that transcripts derived from the maternal seed coat might have contaminated our embryo samples. However, this hypothesis does neither explain our genetic results and reporter gene analyses (Autran et al., 2011) nor other studies (reviewed in Baroux et al. 2009). We demonstrate here that maternal dominance in Ler x Col-0 embryos does not result from a technical artifact but rather may have an interesting biological basis.

The genetic studies and reporter gene analyses we performed suggest that the different findings between the two studies do not arise merely from technical artifacts. Instead, they may have an interesting biological basis, since the two experiments do not only differ in the way the embryos were isolated but in at least two other respects: First, different accessions were used, with Cvi being known for its singular epigenetic configuration involving atypical DNA methylation and transposon insertion patterns (Gazzani et al., 2003; Riddle and Richards, 2002; Saze and Kakutani, 2007), and structural heterochromatin phenotypes reminiscent of a dominant-negative effect on RdDM control (Tessadori et al., 2009). In this respect, the results reported by Nodine and Bartel (2012) would be consistent with our former conclusion that embryos maternally deficient in RdDM components show precocious bi-allelic expression of many genes (Autran et al., 2011). Alternatively, their observations are also consistent with our proposition that the maternal control of paternal expression is expected to become weaker with increasing genetic distance (Autran et al., 2011). Second, while we profiled mRNAs irrespective of their polyadenylation status, Nodine and Bartel (2012) specifically analyzed polyadenylated mRNAs. In animals, cytoplasmic poly(A)-elongation is prevalent as a mechanism for the translational regulation of maternal mRNAs during early development (Benoit et al., 2008; Galili et al., 1988; Lasko 2009). Although data with respect to polyadenylation of plant mRNAs is scarce, the existence of a cytoplasmic polyadenylase (Hunt et al,. 2008) and maternal mRNAs populations with short poly(A)-tail (Grimanelli et al., 2005) makes this a plausible scenario. Given these possible biological differences, future investigations on the mechanisms and natural variation in plant zygotic genome activation promise to shed new light onto this essential phase of the plant life cycle.

At the moment, new hybrid embryonic samples resulting from reciprocal crosses between different accessions, including Ler x Col-0 and Cvi x Col-0 crosses, and amplified with different RNA-seq kits (polyA-based and random priming) await resequencing. The new data will hopefully solve the current controversy concerning parental contribution to the early embryonic transcriptome.

References:

Autran, D., Baroux, C., Raissig, M.T., Lenormand, T., Wittig, M., Grob, S., Steimer, A., Barann, M., Klostermeier, U.C., Leblanc, O., Vielle-Calzada, J.-P., Rosenstiel, P., Grimanelli, D., and Grossniklaus, U. (2011). Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. Cell **145**: 707–719.

Baroux, C., Autran, D., Gillmor, C.S., Grimanelli, D., and Grossniklaus, U. (2009). The Maternal to Zygotic Transition in Animals and Plants. Cold Spring Harb Symp Quant Biol **73**, 89-100.

Benoit, P., Papin, C., Kwak, J.E., Wickens, M., and Simonelig, M. (2008). PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in *Drosophila*. Development **135**: 1969–79.

Galili, G., Kawata, E.E., Smith, L.D., and Larkins, B.A. (1988). Role of the 3'-poly(A) sequence in translational regulation of mRNAs in *Xenopus laevis* oocytes. J Biol Chem **263**: 5764–70.

Gazzani, S., Gendall, A.R., Lister, C., and Dean, C. (2003). Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. Plant Phys **132**: 1107–14.

Grimanelli, D., Perotti, E., Ramirez, J., and Leblanc, O. (2005). Timing of the maternal-to-zygotic transition during early seed development in maize. Plant Cell **17**: 1061–72.

Hunt, A.G., Xu, R., Addepalli, B., Rao, S., Forbes, K.P., Meeks, L.R., Xing, D., Mo, M., Zhao, H., Bandyopadhyay, A., Dampanaboina, L., Marion, A., Von Lanken, C., and Li, Q.Q. (2008). *Arabidopsis* mRNA polyadenylation machinery: comprehensive analysis of protein-protein interactions and gene expression profiling. BMC Genomics **9**: 220.

Lasko, P. (2009). Translational control during early development. Prog Mol Biol Transl Sci **90**: 211–54.

Nodine, M.D. and Bartel, D.P. (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature **482**: 94–97.

Riddle, N.C. and Richards, E.J. (2002). The control of natural variation in cytosine methylation in *Arabidopsis*. Genetics **162**: 355–63.

Saze, H. and Kakutani, T. (2007). Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor *DDM1*. EMBO J **26**: 3641–52.

Tessadori, F., Van Zanten, M., Pavlova, P., Clifton, R., Pontvianne, F., Snoek, L.B., Millenaar, F.F., Schulkes, R.K., Van Driel, R., Voesenek, L.A.C.J., Spillane, C., Pikaard, C.S., Fransz, P., and Peeters, A.J.M. (2009). Phytochrome B and histone deacetylase 6 control light-induced chromatin compaction in *Arabidopsis thaliana*. PLoS Genet **5**: e1000638.

PART 2
CHAPTER 3

GENOMIC IMPRINTING IN THE *ARABIDOPSIS* EMBRYO IS PARTLY REGULATED BY PRC2

NOTE

All of CHAPTER 2.3 will be submitted to PLOS Genetics under the title “Genomic Imprinting in the *Arabidopsis* Embryo Is Partly Regulated by PRC2” with the contributing authors Raissig MT, Baroux C, and Grossniklaus U.

MTR, CB and UG designed the experiments, MTR performed the experiments, MTR and UG analyzed the data, and MTR, CB and UG wrote the manuscript. All supplemental information that will accompany this paper can be found in Appendix A7.

ABSTRACT

Genomic imprinting results in monoallelic gene expression in a parent-of-origin-dependent manner and is regulated by differential epigenetic marking of the parental alleles. It was proposed that genomic imprinting in plants is largely restricted to the endosperm, a tissue nourishing the developing embryo and not contributing to the next generation. In *Arabidopsis*, however, the two imprinted genes *MEDEA* (*MEA*) and *PHERES1* (*PHE1*) are both active in the embryo, but whether the embryonic expression is imprinted or not remains controversial. In addition, one imprinted gene is described in the maize embryo.

We identified several imprinted candidate genes in an allele-specific transcriptome of hybrid *Arabidopsis* embryos and confirmed parent-of-origin-dependent, monoallelic expression for eleven maternally expressed genes (MEGs) and one paternally expressed gene (PEG) in the embryo using allele-specific expression analysis and reporter gene assays. Genetic studies indicate that the *Polycomb* Repressive Complex 2 (PRC2) but not the DNA *METHYLTRANSFERASE1* (*MET1*) is involved in regulating imprinted expression in the embryo. In the seedling, all embryonic MEGs and the PEG are expressed from both parents suggesting that the imprint is erased during late embryogenesis or early vegetative development.

We report here several genes regulated by genomic imprinting in the *Arabidopsis* embryo and show that this epigenetic phenomenon is clearly not a unique feature of the endosperm in both monocots and dicots.

AUTHOR SUMMARY

In most cells nuclear genes are present in two copies, one maternal and one paternal allele. Usually, the two alleles share the same fate regarding their activity with both copies being active or both being silent. An exception to this rule are genes that are regulated by genomic imprinting, where only one allele is expressed and the other one remains silent depending on the parent it was inherited from. The two alleles are equal in terms of their DNA sequence but carry different epigenetic marks and can therefore be distinguished. Genomic imprinting evolved independently in mammals and flowering plants. In mammals, genes regulated by genomic imprinting are expressed in a wide range of tissues including the embryo and the placenta. Yet in plants, genomic imprinting is thought to be restricted to the endosperm, a nutritive tissue in the seed with a function similar to that of the mammalian placenta. Here, we describe the occurrence of genes regulated by genomic imprinting in the embryo of the model plant *Arabidopsis thaliana*. An epigenetic silencing complex, the Polycomb Repressive Complex2 (PRC2), partly regulates genomic imprinting in the embryo. Interestingly, embryonic imprints seem to be erased during late embryo development.

INTRODUCTION

Genes regulated by genomic imprinting are expressed preferentially from one allele in a parent-of-origin-dependent manner. The two alleles do not differ in their DNA sequence, but rather they are differentially marked by epigenetic modifications. In mammals, imprint establishment occurs after meiosis and during gamete development (Reik and Walter, 2001). The parent-of-origin-specific "imprint" is retained after fertilization, such that the alleles can be distinguished and differentially expressed after fertilization. The imprints are erased in the primordial germ cell lineage, which will develop into the gametes, and reestablished according to the sex of the germ line (Barlow, 2011; Feng et al., 2010).

Genomic imprinting evolved both in placental mammals and in flowering plants. While genes can be imprinted in both the embryo and the placenta in mammals, and even in adult tissues (Barlow, 2011; Frost and Moore, 2010), it has been proposed that imprinting in plants is largely restricted to the endosperm, the triploid nourishing tissue that develops upon fertilization of the diploid central cell (Feil and Berger, 2007; Jullien and Berger, 2009). The triploid endosperm does not contribute to the next generation and therefore there is no requirement to reset parental imprints. To date, only two plant genes have been described that display parent-of-origin-dependent, monoallelic expression in both the embryo and the endosperm: *Maternally expressed in embryo1* (*Mee1*) in maize and Os10g05750 in rice, although absence of expression in the gametes and, therefore, *de novo* expression in the fertilization products is only shown for *Mee1* (Jahnke and Scholten, 2009; Luo et al., 2011). In *Arabidopsis thaliana*, the *Polycomb* group gene *MEDEA* (*MEA*) and its target, the MADS-box transcription factor *PHERES1* (*PHE1*), are both imprinted in the endosperm, and show embryonic expression (Köhler et al., 2005; Raissig et al., 2011; Spillane et al., 2007) but it remains controversial whether the embryonic expression itself is imprinted or not (reviewed in Raissig et al., 2011).

The regulation of genomic imprinting in mammals is complex and involves DNA methylation, histone modifications, and non-coding RNAs (Barlow, 2011; Bartolomei and Ferguson-Smith, 2011; Ferguson-Smith, 2011). In *Arabidopsis*, DNA methylation and *Polycomb* Repressive Complex 2 (PRC2)-mediated trimethylation of histone 3 lysine 27 (H3K27me3) are involved in the regulation of some imprinted loci in the endosperm. The maintenance DNA-methyltransferase *METHYLTRANSFERASE1* (*MET1*) and the DNA-glycosylase *DEMETER* (*DME*) act antagonistically to regulate imprinting of *MEA*, *FLOWERING WAGENINGEN* (*FWA*), *FERTILIZATION INDEPENDENT SEED2* (*FIS2*), and *MATERNALLY EXPRESSED PAB C-TERMINAL* (*MPC*) (Choi et al., 2002; Kinoshita et al., 2004; Gehring et al., 2006; Jullien, Kinoshita, et al., 2006; Tiwari et al., 2008). *DME* is preferentially expressed in the central cell and removes DNA-methylation marks on maternal alleles (Choi et al., 2002; Gehring et al., 2006), which, however, might not directly define the imprinting status at all loci (Wöhrmann et al., 2012). In maize, most imprinted genes analyzed in detail are differentially methylated in the endosperm (Jahnke and Scholten, 2009; Gutiérrez-Marcos et al., 2006; Haun et al., 2007; Hermon et al., 2007), which is already established in the gametes for some loci but not for others indicating additional, primary imprinting marks (Jahnke and Scholten, 2009; Gutiérrez-Marcos et al., 2006). Also the regulation of imprinted *MEA* expression by DME and MET1 is indirect and it is not the presence or absence of DNA methylation that distinguishes the two parental alleles (Wöhrmann et al., 2012). However, regulation of imprinted expression of *MEA, PHE1, FORMIN-HOMOLOGUE 5* (*AtFH5*), and some other loci depends additionally or exclusively on the repressive action of PRC2 (Köhler et al., 2005; Baroux et al., 2006; Gerald et al., 2009; Hsieh et al., 2011; Jullien et al., 2006).

Recent studies identified many novel candidate imprinted genes in *Arabidopsis*, maize and rice using systematic genome-wide transcriptome screens on seed tissue (Luo et al., 2011; Hsieh et al., 2011; Gehring et al., 2011; Wolff et al., 2011; McKeown et al., 2011; Waters et al., 2011; Zhang et al., 2011). In *Arabidopsis* alone, the

total number of imprinted genes increased from 12 to more than 300 potentially imprinted genes (Hsieh et al., 2011; Gehring et al., 2011; Wolff et al., 2011; McKeown et al., 2011). Yet, the identified MEGs and PEGs differ dramatically between the studies, which show little overlap even within this species. This might be due to different developmental stages and accessions analyzed, and to different statistical and analytical procedures used to define candidate genes (Gehring et al., 2011; McKeown et al., 2011; Deveale et al., 2012). Nevertheless, future research will elucidate the function and regulation of those candidate genes and add to our current understanding of how imprinting is regulated and how it evolved in plants.

In this study, we show that genomic imprinting is not restricted to the endosperm in *Arabidopsis*, and describe parent-of-origin-dependent, monoallelic expression in the *Arabidopsis* embryo. We identified 80 potentially imprinted genes from a parent-of-origin-specific embryonic transcriptome (Autran et al., 2011) and confirmed eleven MEGs and one early PEG using allele-specific expression analysis of parental transcripts and reporter gene assays. Furthermore, we found that PRC2 is involved in maintaining the imprinted expression pattern at some loci. However, imprinted expression in the embryo requires erasure and resetting of the imprinting marks between the generations. Interestingly, the MEGs and the PEG are expressed from both alleles in the seedling suggesting that the imprint is erased during late embryogenesis or early seedling development.

RESULTS

In-Depth Analysis of the Hybrid Embryonic Transcriptome Reveals Monoparentally Expressed Genes

To study the global parental contributions to the embryonic transcriptome and its regulation, embryonic samples were previously generated from hybrid embryos at the 2-4 cell and globular stage (Autran et al., 2011). The hybrid embryos were derived either from a cross between Landsberg *erecta* (Ler) and Columbia-0 (Col-0) or from a cross between the *kryptonite* (*kyp*) mutant (Ler) and Col-0. Subsequent high-throughput sequencing of the generated cDNA libraries allowed the identification of allele-specific transcripts based on single nucleotide polymorphisms (SNPs) between the accessions (Autran et al., 2011). For this study, we identifed potentially imprinted transcripts in the embryo, for which only one parental allele was sequenced in all samples, in the 2-4 cell samples only or in the globular samples only (Autran et al., 2011; for details see Material and Methods). We only filtered genes that were not deregulated in the *kyp/KYP* x Col-0 sample, assuming that *KYP* is not a major regulator of genomic imprinting, but rather regulates the parental contributions at the genome-wide level. This procedure yielded 50 potential maternally expressed genes (MEGs) and 30 potential paternally expressed genes (PEGs) in the *Arabidopsis* embryo (Table S2-3-1). A recent study analyzing a hybrid embryonic transcriptome in *Arabidopsis* (Nodine and Bartel, 2012) confirmed the presence of monoallelic gene expression in the *Arabidopsis* embryo. The authors describe 77 maternally and 45 paternally contributed transcripts in at least one embryonic stage tested. Finally, we chose 18 MEG candidates and six PEG candidates that are highly expressed in our embryonic libraries (Autran et al., 2011) and that are absent from the gametes (Wuest et al., 2010; Borges et al., 2008), suggesting *de novo* expression in the embryo, and analyzed them in detail (i.e. allele-specific expression and reporter gene analysis; Table S2-3-1).

Monoallelic Gene Expression in the 2-4 Cell and the Globular Embryo

To confirm the previously identified MEGs and PEGs, we produced new embryonic cDNA libraries by crossing the Col-0 and the Ler accessions reciprocally. Embryos were isolated at the 2-4 cell embryo stage (~2.5 days after pollination (DAP)) and at the globular embryo stage (~4 DAP). We sampled two biological replicates of each cross and stage (8 samples in total), extracted total RNA and amplified a cDNA library (see Material and Methods). The cDNA samples from hybrid embryos were subsequently used to amplify the polymorphic, SNP-containing sequence of potentially imprinted transcripts by RT-PCR, and products were assessed for parent-of-origin by Sanger sequencing. As a control, we performed allele-specific expression analysis on a polymorphic gene that is expressed from both parental alleles (*AT1G02780*, *EMBRYO DEFECTIVE 2386*, Autran et al., 2011). We readily detected both parental nucleotides at the polymorphic site in all the samples analyzed (Figure S2-3-1A), confirming that this method is suitable to detect biallelic gene expression. Importantly, we verified that all the assays used in this study amplify both parental alleles with equal efficiency. To this aim we performed PCR and Sanger sequencing on genomic DNA from reciprocal F1 hybrid seedlings (Col-0 x Ler and Ler x Col-0, respectively). All assays used in this study amplified both alleles with equal efficiency and thus do not introduce a technical bias towards one allele (Figure S2-3-1B and Figure S2-3-2).

We then performed allele-specific expression analysis on the previously selected 18 candidate MEGs and six candidate PEGs (Table S2-3-1). For eleven of the 18 candidate MEGs we could sequence only the maternal allele in all crosses, stages and replicates analyzed (*AT1G29660, AT1G72260, AT2G47115, AT5G62210, AT3G20520, AT2G17710, AT3G21500, AT2G01520, AT1G20680, AT5G51950, AT1G29050;* Figure 2-3-1A - F and Figure S2-3-3A - E). Because 9 of the 11 genes showed no detectable levels in the egg transcriptome, our result strongly suggests that they may be regulated by genomic imprinting in the *Arabidopsis* embryo (Borges et

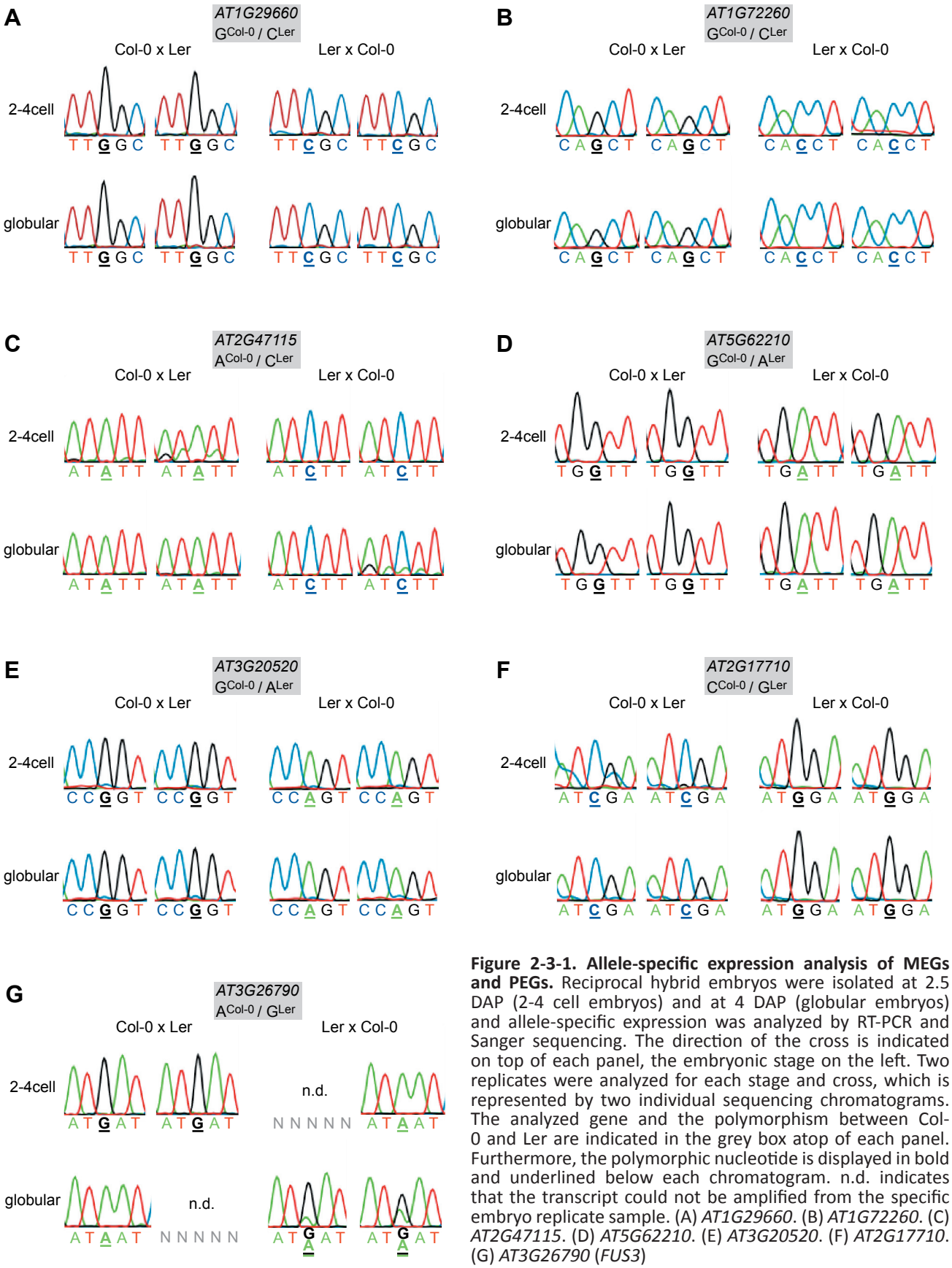


Figure 2-3-1. Allele-specific expression analysis of MEGs and PEGs. Reciprocal hybrid embryos were isolated at 2.5 DAP (2-4 cell embryos) and at 4 DAP (globular embryos) and allele-specific expression was analyzed by RT-PCR and Sanger sequencing. The direction of the cross is indicated on top of each panel, the embryonic stage on the left. Two replicates were analyzed for each stage and cross, which is represented by two individual sequencing chromatograms. The analyzed gene and the polymorphism between Col-0 and Ler are indicated in the grey box atop of each panel. Furthermore, the polymorphic nucleotide is displayed in bold and underlined below each chromatogram. n.d. indicates that the transcript could not be amplified from the specific embryo replicate sample. (A) *AT1G29660*. (B) *AT1G72260*. (C) *AT2G47115*. (D) *AT5G62210*. (E) *AT3G20520*. (F) *AT2G17710*. (G) *AT3G26790* (*FUS3*)

al., 2008; Wuest et al., 2010, Table S2-3-1). Two of the 18 candidate MEGs showed biallelic expression in one 2-4 cell replicate sample and were therefore excluded from further analyses (*AT3G44260* and *AT5G52060*, Figure S2-3-3F and S2-3-3G). However, the maternal signal was much higher in these replicates, and both genes showed complete monoallelic expression at the globular stage, suggesting that they may also be regulated by genomic imprinting. From the remaining five MEGs, four genes could not be amplified at all from the embryonic cDNA libraries and one gene was lost when a highly stringent washing procedure was applied before RNA extraction and amplification (see below, Table S2-3-1) suggesting that they are not or only weakly expressed in the embryo.

Of the six analyzed candidate PEGs, only one PEG (*FUSCA 3* (*FUS3, AT3G26790*)) showed consistent monoallelic expression at the 2-4 cell stage, but not at the globular stage, at which expression was either maternal or biallelic, depending on the direction of the cross (Figure 2-3-1G). In two replicates, *FUS3* could not be detected at all, which might indicate that *FUS3* is expressed at low level in the embryo; even below detection level in some samples. Two other candidate PEGs were found to be expressed from both parents (*AT2G20160* (*MEIDOS*) and *AT1G63260*; Figure S2-3-4), while the three remaining candidate PEGs could not be detected at all in our embryonic cDNA libraries (Table S2-3-1).

In summary, we could confirm eleven MEGs and one PEG that show parent-of-origin-dependent monoallelic gene expression in the embryo. For nine MEGs and the PEG no transcripts were found in the gametes suggesting *de novo* expression and, therefore, regulation by genomic imprinting in the embryo (Wuest et al., 2010; Borges et al., 2008). One MEG, *AT1G72260*, is expressed in the gametes already, and another MEG, *AT2G47115*, is not represented on the microarrays and no statement can be made about their active expression in the embryo (Wuest et al., 2010; Borges et al., 2008). Nevertheless, genomic imprinting in the embryo is more widespread than commonly thought and imprinted expression of the maize gene *mee1* is not an exceptional case.

The Monoallelic Expression Pattern Is Not Due to Sporophytic or Endosperm Contamination

The embryo is surrounded by the triploid endosperm, and both embryo and endosperm are embedded in the maternal, sporophytic seed coat. Therefore, isolating embryos devoid of debris from surrounding sporophytic tissues and careful control of maternal tissue contamination is an important issue (Nodine and Bartel, 2012). While the PEG cannot derive from seed coat contamination, substantial contamination with maternal sporophytic tissue could explain the observed maternal expression patterns of the confirmed MEGs. Our initial samples were prepared from embryos washed one time and although all samples were devoid of visible debris at collection we produced two additional embryonic cDNA libraries to rule out the possibility of sporophytic contamination: following reciprocal crosses between Col-0 and Ler, we isolated 2-4 cell embryos, but washed them six times (6x) instead of one time (1x) only, removing all possible non-embryonic transcripts.

First, we assessed the quality of the 6x washed embryonic cDNA libraries compared to the 1x washed libraries by performing RT-PCR using primers amplifying *ACTIN 11* (*ACT11*) and *WUSCHEL-RELATED HOMEOBOX 9* (*WOX9*), an embryo-specific gene (Figure 2-3-2A). We amplified both control genes in the 1x and the 6x washed embryonic cDNA libraries derived from the Col-0 x Ler cross (Figure 2-3-2A). In contrast, we could hardly detect *ACT11* and only weakly amplify *WOX9* in the 6x washed Ler x Col-0 library (Figure 2-3-2A). This indicates a lower cDNA library quality, likely due to RNA degradation during the washes. Nevertheless, we could readily detect the transcript of seven MEGs (*AT1G29660, AT1G72260, AT2G47115, AT5G62210, AT3G20520, AT2G17710, AT3G21500)* in both 6x washed 2-4 cell embryo samples and confirmed their monoallelic expression pattern with Sanger sequencing analysis (Figure 2-3-2B, Figure S2-3-5A). For the other four MEGs (*AT2G01520, AT1G20680, AT5G51950, AT1G29050*) we could only detect expression in the 6x washed Col-0 x Ler embryo sample, but not in the 6x washed Ler x Col-0 embryo sample, likely due to lower cDNA quality of this particular library. Nevertheless, Sanger sequencing analysis confirmed that those four MEGs show monoallelic expression

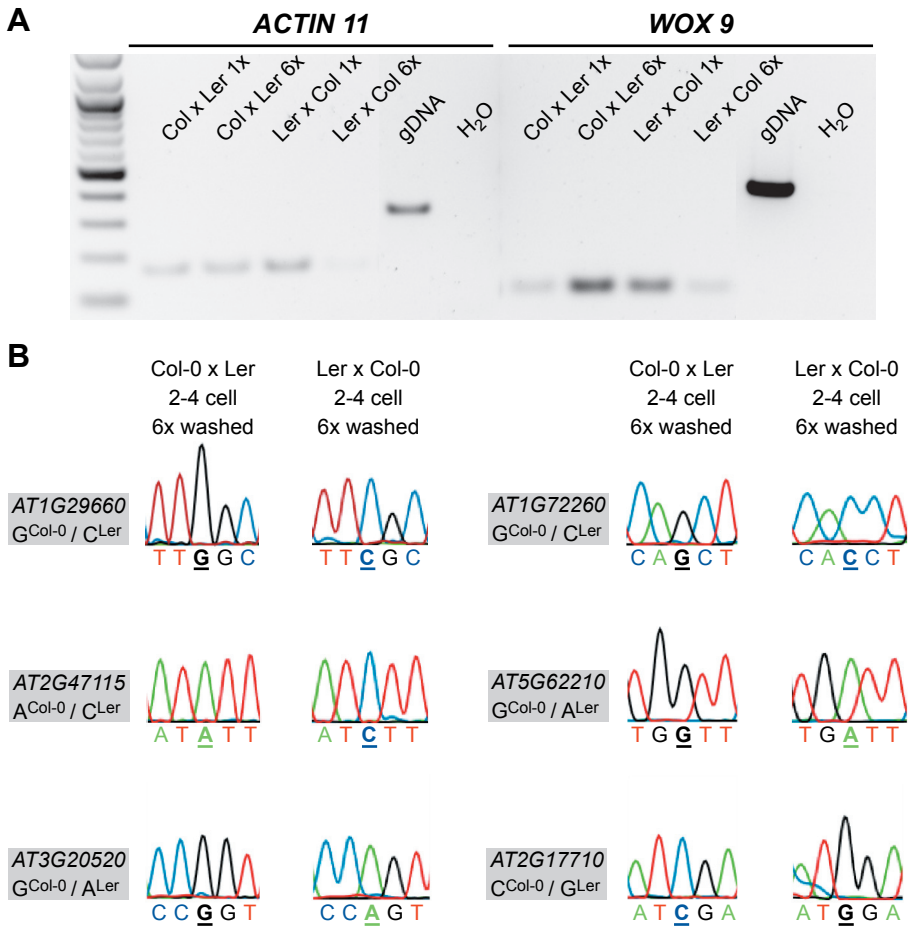


Figure 2-3-2. Allele-specific expression analysis of extensively washed embryonic control samples. (A) RT-PCR amplifying *ACT11* and *WOX9*, an embryo-specific gene, from 1x washed and 6x washed reciprocal, embryonic cDNA libraries (2-4 cell stage) using 32 PCR cycles. Genomic DNA was used as positive and water as negative control. (B) Allele-specific expression analysis of *AT1G29660*, *AT1G72260*, *AT2G47115*, *AT5G62210*, *AT3G20520*, and *AT2G17710* in the 6x washed embryonic cDNA libraries. The analyzed gene and the polymorphism between Col-0 and Ler are indicated in the grey box and the direction of the cross on top of the panel. The polymorphic nucleotide is displayed in bold and underlined below each chromatogram.

in the 6x washed Col-0 x Ler sample (Figure S2-3-5A) and we therefore classify them as partially confirmed MEGs. Only one candidate gene that seemed to be imprinted in the 1x washed libraries could not be detected anymore in both 6x washed embryo samples, suggesting that this transcript is either low abundant in the embryo or that it indeed derived from maternal sporophytic and/or endosperm contamination (*AT4G11960*, Figure S2-3-5C). In fact, the loss of only one of twelve candidate MEGs after extensive washing suggests that our previous washing regime is indeed sufficient to remove non-embryonic transcripts.

In conclusion, sequencing analysis of extensively washed, reciprocal 2-4 cell embryonic cDNA libraries fully confirmed seven MEGs and partially confirmed four MEGs as imprinted genes in the embryo, and shows that their monoallelic expression is not caused by seed coat or endosperm contamination.

MEG and PEG Reporter Lines Show Imprinted Expression in the Embryo

In order to demonstrate parent-of-origin-dependent, monoallelic expression in the embryo with an independent assay, we cloned the promoter of seven MEGs (*AT1G29660*, *AT1G72260*, *AT2G47115*, *AT5G62210*, *AT3G20520*, *AT2G17710*, *AT3G21500*) and the single identified PEG (*AT3G26790*) as transcriptional fusions with the *uidA* reporter gene encoding ß-Glucuronidase (GUS; Jefferson et al., 1987). We screened 24 independent T1 lines for all 8 constructs for expression in the seed. Except for the PEG reporter *pFUS3::GUS*, all MEG reporters exhibited fairly strong staining in the seed coat, making embryo expression analysis on whole seeds impossible (Figure S2-3-6). To assess whether the gene-of-interest is indeed expressed in the embryo as indicated by our previous analyses, we isolated self-fertilized embryos at early and late stages of two to three independent lines for each MEG construct in the T1 generation and stained them for GUS activity on slides. Whereas 6 lines showed strong expression in the embryo (Figure S2-3-7), one line (*pAT3G21500::GUS*) showed only very weak and hardly detectable GUS staining in the embryo (Figure S2-3-8) and was therefore not used for further analysis.

To assess whether the promoter-GUS reporters are imprinted in the embryo, we performed reciprocal

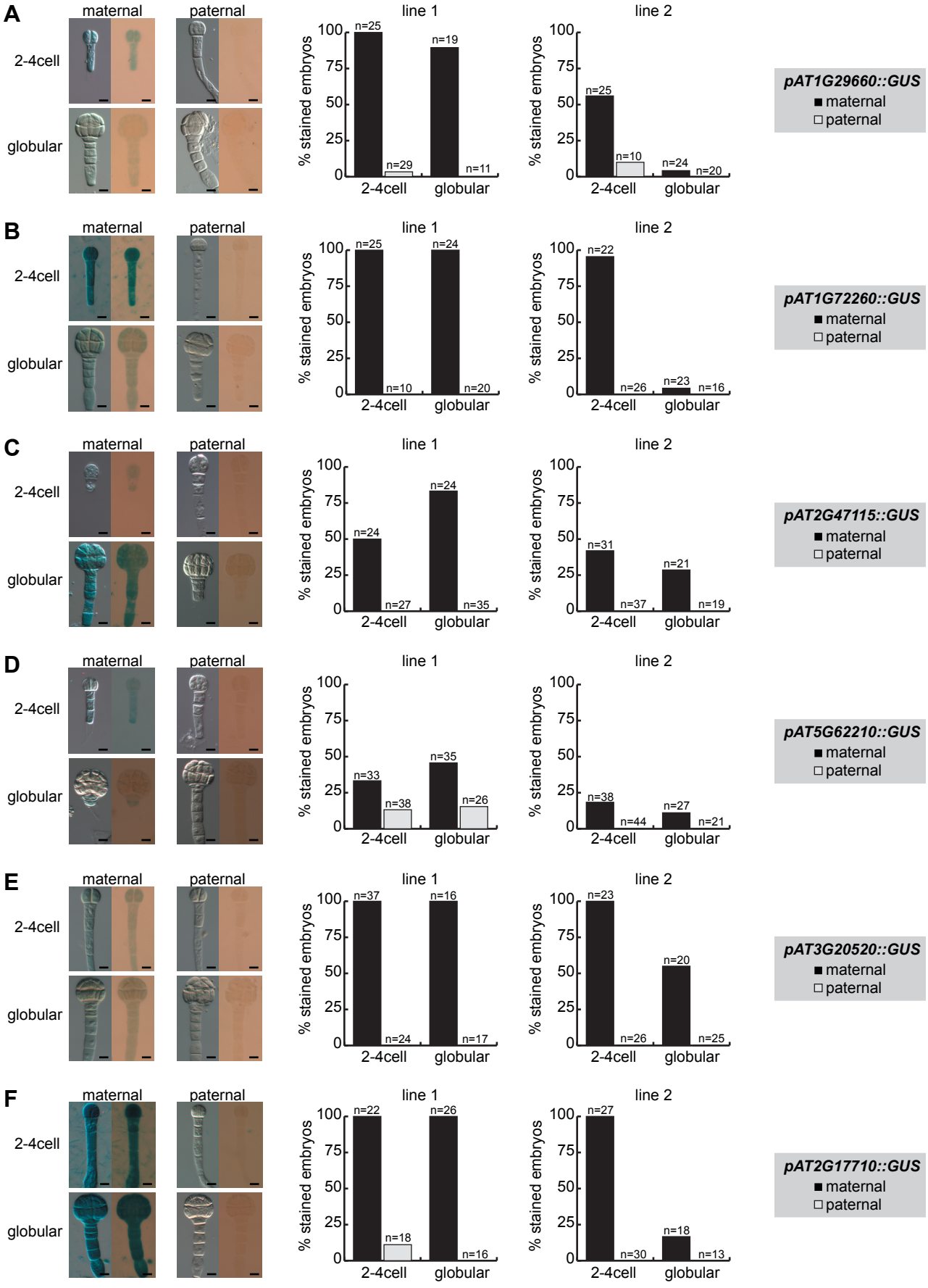
crosses between two independent reporter lines of each of the six strong MEG constructs (T2 generation) with wild-type plants (Col-0). F1 embryos were isolated at 2.5 DAP (~2-4-cell embryos) and 4 DAP (~globular embryos), and stained on slides for 4 days before analyzing them for GUS expression. For the MEG reporters, we expected to see the GUS signal only in the embryos that inherited the reporter gene maternally and not in the embryos that received the reporter gene from the pollen donor. We found that three of the six GUS-reporter lines are fully imprinted showing exclusive maternal expression (*pAT1G72260::GUS, pAT2G47115::GUS, pAT3G20520::GUS,* Figure 2-3-3B, 2-3-3C, and 2-3-3E), while the remaining three GUS-reporter lines show a very strong bias towards maternal expression *(pAT1G29660::GUS, pAT5G62210::GUS, pAT2G17710::GUS,* Figure 2-3-3A, 2-3-3D, and 2-3-3F). The PEG reporter line *pFUS3::GUS* shows very strong and embryo-specific expression starting from the 8 cell embryo stage both in whole mount seed staining assays (Figure S2-3-6H) and after embryo isolation (Figure S2-3-9C). Since GUS expression for this gene seems to be very specific to the embryo (Figure S2-3-6H), no embryo isolation was necessary after reciprocal crosses to assess whether *pFUS3::GUS* shows imprinted expression. Yet, in contrast to the MEG reporter lines, we detected embryonic GUS expression no matter from which parent the reporter was inherited. This suggests that the upstream regulatory region of *FUS3* is not sufficient to confer imprinted paternal expression in early *Arabidopsis* embryos. However, *pFUS3::GUS* activity was first detected at 3 DAP corresponding to the (4-)8 cell stage. Thus, the level of gene expression at earlier stages, where we actually detected exclusively paternal expression using allele-specific expression analysis (Figure 2-3-1G), might be below detection level in this assay.

In conclusion, all the MEG and PEG reporter lines cloned and analyzed are expressed in the embryo (Figure 2-3-3, Figure S2-3-7, Figure S2-3-9). Moreover, all MEG reporter lines are either fully imprinted or show a strong bias for maternal expression (Figure 2-3-3). The upstream regulatory sequences that were cloned are, thus, sufficient to confer imprinted expression during early stages of embryogenesis. On the other hand, a few loci, such as *FUS3,* might require additional regulatory elements for imprinting. Finally, all MEG reporter lines are expressed in the seed coat (Figure S2-3-6) and the embryo (Figure S2-3-7), but are clearly regulated by genomic imprinting in the embryo (Figure 2-3-3).

PRC2 but not MET1 Is Involved in Regulating Genomic Imprinting in the *Arabidopsis* Embryo

In order to investigate how genomic imprinting is regulated in the embryo, we crossed mutants affecting imprinting regulators to wild-type parents of a distinct accession. DNA-methylation and histone modification, in particular H3K27me3 mediated by the PRC2, have both been shown to regulate genomic imprinting (for review see Raissig et al., 2011). Therefore, we crossed the *fertilization-independent endosperm* (*fie*) mutant, in which PRC2-mediated repression is fully abolished, reciprocally to wild-type plants, and used the *methyltransferase 1* (*met1-3*) mutant, disrupting DNA-methylation maintenance in the CG-context, as donor to pollinate wild-type plants. *MET1* was thus far only implicated as paternal repressor of imprinted loci, whereas PRC2 regulates MEGs (eg. *MEA*) and PEGs (eg. *PHE1,* for review see Raissig et al., 2011). Therefore, we crossed *fie* mutants reciprocally but the *met1-3* mutant only as father. As before, we isolated embryos from the resulting F1 hybrid seeds and proceeded with

Figure 2-3-3. Parent-of-origin-dependent expression of MEG reporter lines in isolated embryos. The MEG reporter lines were reciprocally crossed to wild-type Col-0 plants and embryos were isolated at 2.5 DAP (2-4 cell embryos) and at 4 DAP (globular embryos) prior to GUS staining. Embryos were stained on slides for 4 days at 37°C and then analyzed for GUS expression using bright-field microscopy. For each line, two independent transgene insertions (line 1, line 2) were analyzed and quantified for maternal (black columns, maternally inherited reporter gene) and paternal expression (grey columns, paternally inherited reporter gene) and are displayed separately (middle panel and right panel). Embryo pictures of line 1 are shown on the left, always showing a DIC picture and a bright-field picture of each stage and direction of cross. The embryonic stages, maternal or paternal GUS reporter expression and the analyzed reporter line are indicated and the numbers of the quantifications are shown above each column. Scale bar = 10 µm. (A) p*AT1G29660::GUS*. (B) *pAT1G72260::GUS*. (C) *pAT2G47115::GUS*. (D) *pAT5G62210::GUS*. (E) *pAT3G20520::GUS*. (F) *pAT2G17710::GUS*.


A
maternal
paternal
2-4cell
globular
line 1
line 2
% stained embryos
100
75
50
25
0
n=25
n=29
n=19
n=11
n=25
n=10
n=24
n=20
2-4cell
globular
pAT1G29660::GUS
maternal
paternal
B
n=25
n=10
n=24
n=20
n=22
n=26
n=23
n=16
pAT1G72260::GUS
C
n=24
n=27
n=24
n=35
n=31
n=37
n=21
n=19
pAT2G47115::GUS
D
n=33
n=38
n=35
n=26
n=38
n=44
n=27
n=21
pAT5G62210::GUS
E
n=37
n=24
n=16
n=17
n=23
n=26
n=20
n=25
pAT3G20520::GUS
F
n=22
n=18
n=26
n=16
n=27
n=30
n=18
n=13
pAT2G17710::GUS

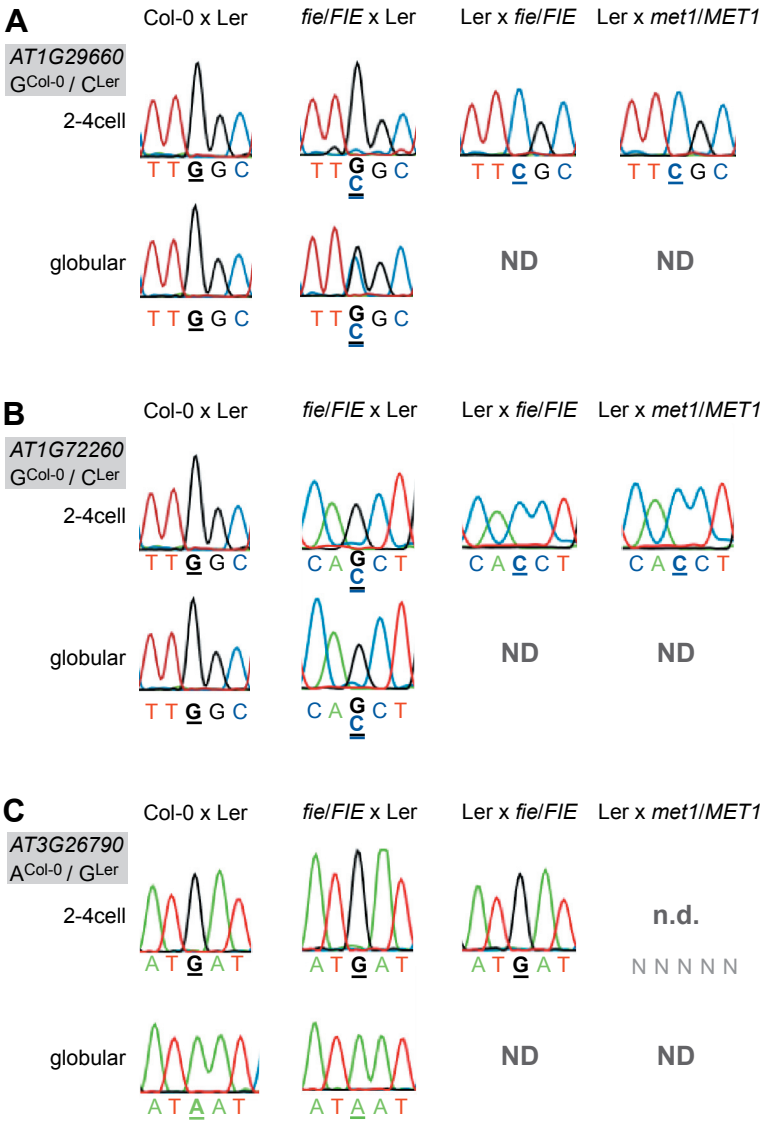


Figure 2-3-4. Effect of PRC2 and *MET1* function on imprinted expression in the embryo. Mutant embryonic samples were generated and the confirmed MEGs and the PEG were analyzed for derepression of the silent allele. Heterozygous *fie* mutants were crossed maternally and paternally and heterozygous *met1-3* mutants were crossed paternally as indicated above the chromatograms. To simplify the reading of the graph, we present again the wild-type situation of one replicate of the Col-0 x Ler cross from Figure 1 at both embryonic stages tested. Embryos were isolated at 2.5 DAP (2-4 cell embryos) and at 4 DAP (globular embryos, only for the cross *fie*/*FIE* x Ler). The embryonic stage is indicated on the left, the analyzed gene and the polymorphism between the mutant (all in Col-0 background) and the wild-type allele (Ler) is shown in the grey box beside each panel. Furthermore, the polymorphic nucleotide is displayed in bold and underlined below each chromatogram. ND indicates that the sample was not available; n.d. indicates that the transcript was not detected. (A) *AT1G29660*. (B) *AT1G72260*. (C) *FUS3* (*AT3G26790*).

RNA extraction and library amplification, creating mutant embryonic cDNA libraries (i.e. *fie*/*FIE* x Ler, Ler x *fie*/*FIE*, and Ler x *met1-3*/*MET1*).

We tested the allele-specific expression pattern of the eleven MEGs and the PEG in the Ler x *met1-3*/*MET1* mutant library (2-4cell stage) and found that all MEGs were still monoallelically expressed (Figure 2-3-4 and Figure S2-3-10). Thus, in contrast to some of the well-studied maternally expressed imprinted loci in the *Arabidopsis* endosperm (ie. *FIS2*, *FWA*), disrupting paternal DNA-methylation maintenance does not appear to affect the imprinted expression of the embryonic MEGs at all (Figure 2-3-4A and 2-3-4B and Figure S2-3-10). Yet, the PEG could not be detected anymore in the Ler x *met1-3*/*MET1* mutant cDNA library, indicating an involvement of paternal *MET1* in activating the paternal *FUS3* allele (Figure 2-3-4C).

However, disrupting PRC2 function by crossing *fie* mutants maternally or paternally did have an effect on two MEGs and on the PEG. A maternal *fie* mutation was able to derepress the paternal alleles of *AT1G29660* and *AT1G72260* (Figure 2-3-4A and 2-3-4B). The alleles were slightly derepressed at the 2-4 cell stage in both cases, but were differently affected by the *fie* mutation at the globular stage. While *AT1G29660* was fully derepressed and biallelically expressed (Figure 2-3-4A), the paternal allele of *AT1G72260* retained a low expression level (Figure 2-3-4B). Interestingly, a paternal *fie* mutation induced maternal expression of the PEG *FUS3*, while abolishing its paternal expression (Ler x *fie*/*FIE*, 2-4cell stage, Figure 2-3-4C). Our result suggests that paternal *FIE* activity is required to activate the paternal *FUS3* allele and that paternal FUS3 may negatively control the maternal *FUS3* allele after fertilization.

In conclusion, the PRC2 seems to be involved in regulating genomic imprinting in the embryo. Maternal PRC2 activity maintains the repression of the silent paternal allele of two MEGs after fertilization, while paternal PRC2 function together with paternal MET1 is somehow implicated in the activation of the paternal allele of

the PEG. In contrast, paternal MET1 activity does not seem to play a role in regulating genomic imprinting of the eleven MEGs in the embryo. Our result indicates that there must be additional factors regulating genomic imprinting in the embryo.

Disrupting Embryonic MEGs or PEGs Has No Effect on Seed Viability or Early Embryogenesis

Some of the imprinted genes in the endosperm, like *MEA* and *FIS2*, show parent-of-origin-dependent seed abortion when mutated (Chaudhury et al., 1997; Grossniklaus et al., 1998). To reveal a potential role of the confirmed embryonic MEGs and the PEG during embryogenesis and seed development, we analyzed T-DNA insertions (Table S2-3-2). We assessed seed viability by dissecting siliques and analyzing seed set of 16 to 24 individuals of a genotyped segregating population. None of the analyzed T-DNA insertion lines showed reduced seed set (Table S2-3-2). This suggests that the MEGs and PEGs we identified have a more subtle role during embryogenesis or do not show an effect on seed development due to gene redundancy or due to an incomplete disruption of the targeted gene. Furthermore, we tried to assess more subtle effects by dissecting and clearing seeds of heterozygous mutant individuals followed by morphological analysis of early embryogenesis. However, we could not observe any obvious patterning defects or other developmental aberrations in the lines analyzed (Table S2-3-2). Interestingly, homozygous *fus3* embryos show a phenotype late in seed development, namely a prolonged cell division phase in the embryo throughout seed maturation (Raz et al., 2001). Yet, the late occurrence of this phenotype is unlikely to be caused by the early imprinted state of *FUS3*, and the recessive nature of the phenotype fits well with our observation that *FUS3* shows a biallelic expression late in embryogenesis.

Taken together, we did not identify any fertility or embryo patterning phenotypes when analyzing T-DNA insertion lines disrupting the imprinted embryonic genes we identified.

Most MEGs and PEGs Are Contributed in a Parent-of-Origin-Dependent Manner at Later Stages and in Different Ecotypes

To determine whether our embryonic MEGs and PEG are (i) indeed expressed in other embryonic samples and (ii) monoallelically contributed in other *Arabidopsis* accessions and at different embryonic stages, we compared our data with the results of three recent studies. Xiang and colleagues isolated embryos from Col-0 wild-type plants manually, from zygote up to mature embryos, and performed transcriptome analysis using microarrays (Xiang et al., 2011). Eight of our MEGs and the PEG are expressed clearly above their background level at the 4 cell embryo and the globular embryo stage, whereas three MEGs are just at background level, which is considered not or lowly expressed (Table 2-3-1, Table S2-3-3). In contrast, these three MEGs are clearly expressed in young hybrid embryos from reciprocal crosses between the accession Col-0 and Cape Verde Islands (Cvi; Nodine and Bartel, 2012; Table 2-3-1, Table S2-3-4). This finding suggests, that either the microarray technique is not sensitive enough to detect low expression levels, or that these three MEGs are stronger expressed in hybrids than in self-fertilized embryos.

Furthermore, we compared parent-of-origin-dependent expression of our embryonic MEGs and PEG to the parent-of-origin-dependent expression of those in (i) early Col-0 x Cvi embryos (different accession, similar stage, Nodine and Bartel, 2012) and (ii) late torpedo-stage Col-0 x Ler embryos (same accessions, but later stage, Gehring et al., 2011). Interestingly, eight MEGs and the PEG show a mono-allelic or at least a clearly biased parent-of-origin-dependent expression in early Col-0 x Cvi embryos confirming our study (Table 2-3-1, Table S2-3-4). This suggests that genomic imprinting of those loci might be conserved between accessions. Two MEGs are biallelically expressed and one is low expressed and no SNPs were covered by reads, making a parent-of-origin analysis impossible (Table 2-3-1, Table S2-3-4). Yet, when looking at the same accessions but later in development (torpedo stage) we found that eight MEGs are still contributed maternally or with a maternal bias (although most

Table 1. Embryonic and parent-of-origin expression of the confirmed MEGs and the PEG in other studies

	Embryo expression[1]		Parent-of-origin of transcripts	
Gene	4 cell	globular	1-32 cell[2]	torpedo[3]
AT1G29660	YES	YES	bi-allelic	Ler bias
AT1G72260	YES	YES	maternal	maternal
AT2G47115	NO	NO	maternal	not expressed
AT5G62210	YES	YES	maternal	bi-allelic
AT3G20520	NO	NO	maternal	maternal bias
AT2G17710	YES	YES	bi-allelic	maternal bias
AT3G21500	YES	YES	maternal	maternal
AT2G01520	YES	YES	maternal bias	maternal
AT1G20680	NO	NO	maternal	maternal
AT5G51950	YES	YES	maternal bias	maternal
AT1G29050	YES	YES	no SNPs covered	maternal bias
AT3G26790	YES	YES	early PEG	bi-allelic

[1]: The expression values from (Xiang et al., 2011) at the 4cell and the globular embryo of the 11 MEGs and the PEG are displayed. Detailed expression scores can be found in Table S3.
[2]: The parent-of-origin-dependent expression of the 11 MEGs and the PEG in early embryos (1-32cell) of reciprocal crosses between Col-0 and Cvi (Nodine and Bartel, 2012) is shown. Detailed read counts can be found in Table S4.
[3]: The parent-of-origin-dependent expression of the 11 MEGs and the PEG in late embryos (torpedo stage) of reciprocal crosses between Col-0 and Ler (Gehring et al., 2011) is shown. Detailed read counts can be found in Table S4.

were covered by few reads only). One MEG shows a Ler bias, one MEG and *FUS3* are biallelically expressed and one MEG is not expressed anymore (Table 2-3-1, Table S2-3-4). This indicates, that at some loci imprinted expression is reset already during embryogenesis, whereas at other loci the resetting happens later during or even after embryogenesis. This analysis shows that the majority of the imprinted loci in the embryo is imprinted in different accessions (Nodine and Bartel, 2012) and maintains a parent-of-origin-dependent expression also later during embryo development (Gehring et al., 2011).

Embryonic Imprints Are Erased in the Seedling

Parent-of-origin-dependent expression seems to be maintained until late stages of embryogenesis, at least for some loci analyzed. Eventually the imprint has to be erased, the very latest during sporogenesis and before gametogenesis. To address this question we reciprocally crossed Col-0 and Ler, grew F1 seedlings up to the 4-leaf stage (8 days after sowing) and produced hybrid F1 seedling cDNA libraries. We then performed allele-specific expression analysis using the MEG and PEG assays. 9 of 11 MEGs and *FUS3* could be amplified from the seedling library and are thus expressed in the seedling. The two genes that were not amplified (*AT2G47115*, *AT3G21500*) are either not expressed in the seedling or below the detection level of our assays. Sanger sequencing revealed that all MEGs and *FUS3* are expressed from both parents in F1 hybrid seedlings (Figure 2-3-5). This suggests that the imprint is erased late during embryogenesis or early during vegetative development, but long before flowering and the initiation of reproductive development,

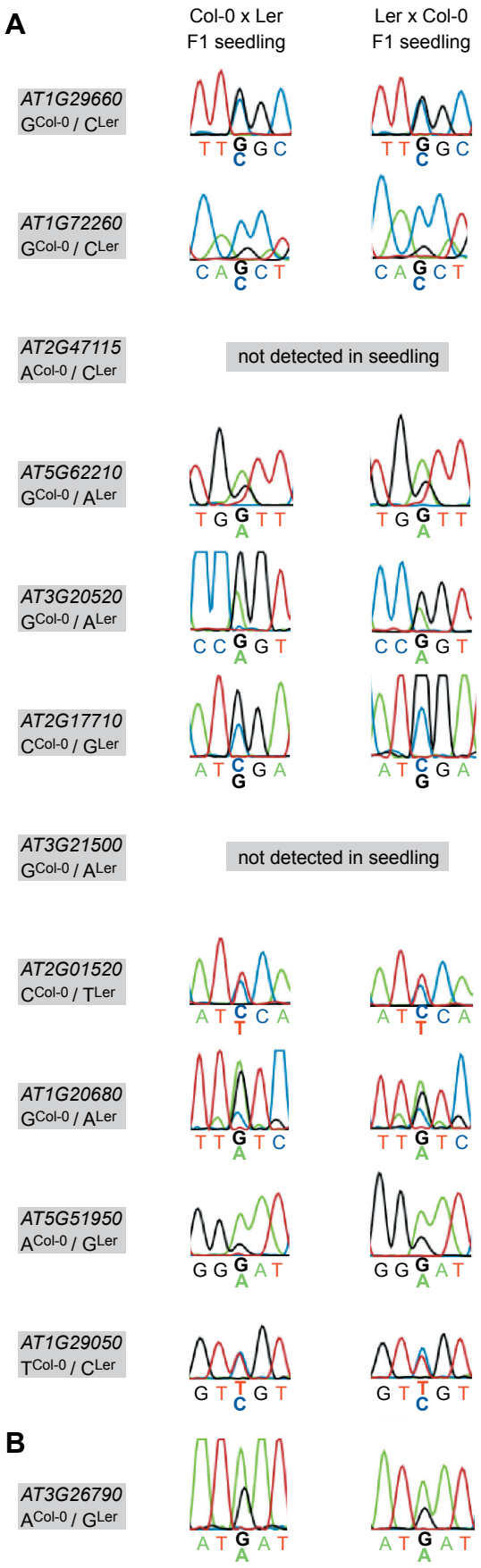


Figure 2-3-5. Allele-specific expression analysis of confirmed MEGs and the PEG in hybrid F1 seedlings. The allele-specific expression of the eleven confirmed MEGs (A) and the confirmed PEG (B) was assessed in reciprocal F1 hybrid seedling cDNA libraries (8 days after sowing). Nine MEGs and the PEG show biallelic expression in the seedling and two MEGs were not detected in the seedling samples, as indicated (*AT2G47115*, *AT3G21500*). The analyzed gene and the polymorphism between Col-0 and Ler are specified in the grey box beside each panel.

DISCUSSION

Parent-of-Origin-Dependent, Monoallelic Gene Expression Is not Restricted to the Endosperm in *Arabidopsis*

Recently, two studies independently identified monoallelically-derived transcripts in early *Arabidopsis* embryos while assessing the genome-wide parental contribution to plant embryogenesis (Autran et al., 2011; Nodine and Bartel, 2012). Analysis of reciprocal Col-0 x Cvi F1 embryos identified more than 100 potentially imprinted or maternally/paternally deposited transcripts in the *Arabidopsis* embryo (Nodine and Bartel, 2012), whereas we identified 50 potential MEGs and 30 potential PEGs by assessing the allele-specific transcriptome of Ler x Col-0 embryos (Autran et al., 2011). We focused on 18 MEG and six PEG candidates that showed strong expression in the embryo but no expression in the gametes. We could confirm eleven MEGs expressed at the 2-4 cell and globular embryo stage and one PEG at 2-4 cell stage using RT-PCR and Sanger sequencing on replicated and reciprocal hybrid, embryonic cDNA samples (Table 2-3-2, Figure 2-3-1, Figure S2-3-3). In addition, reporter gene analysis of seven MEGs and one PEG confirmed embryonic expression and six MEG reporter lines were fully imprinted or expressed with a strong maternal bias in the embryo (Table 2-3-2, Figure 2-3-3, Figure S2-3-7). Furthermore, absence of most MEG or PEG transcripts in the gametes prior to fertilization strongly suggests *de novo* expression of the genes in the embryo (Table 2-3-2, Table S2-3-1, Wuest et al., 2010; Borges et al., 2008). Importantly, the MEGs were not only expressed in the embryo, but also in the

surrounding seed coat (Figure S2-3-6). Therefore, examining potential maternal sporophytic contamination was essential to our study. To do so, we produced two extensively washed (6x), reciprocal 2-4 cell embryonic cDNA libraries that should be completely devoid of all potential contamination. Allele-specific expression analysis in those samples revealed pure monoallelic expression of the 11 MEGs and the PEG in the embryo, affirming the imprinted expression pattern (Figure 2-3-2, Figure S2-3-5). In addition, we isolated embryos derived from reciprocal crosses between MEG reporter lines and wild-type plants prior to GUS staining. Thus, the observed GUS signal is embryo-specific and cannot be due to diffusion or transport from the seed coat.

The first gene described to be imprinted in plant embryos was the maize gene *Mee1* (Jahnke and Scholten, 2009). Other recently published studies performed genome-wide analysis of parent-of-origin allelic expression in endosperm and embryos. Hsieh and colleagues identify 37 MEGs and one PEG in the *Arabidopsis* embryo 7-8 DAP, but discarded them due to possible contamination by maternal tissues (Hsieh et al., 2011). Gehring and coworkers found 17 MEGs and one PEG in *Arabidopsis* embryos 6 DAP (torpedo stage) but none was further analyzed (Gehring et al., 2011). In rice, a similar study identified seven putative MEGs in the embryo and confirmed one, Os10g05750, using RT-PCR and Sanger sequencing but its expression status in the gametes was not assessed (Luo et al., 2011). Furthermore, Waters and colleagues identified 29 MEGs and 9 PEGs from a dataset of maize embryos 14 DAP, but did not follow up these findings with additional experiments (Waters et al., 2011). In all studies, many of the embryonic MEGs and PEGs are also expressed and imprinted in the endosperm, which leads the authors to attribute their pattern of expression to endosperm contamination, and do not consider it a real, intrinsic and embryo-specific feature. Similarly, the confirmed eleven MEGs of this study are not only expressed in the embryo, but also in the seed coat, the seedling and, likely, in the endosperm (Figure S2-3-6). *AT1G20680* even shows imprinted expression in the endosperm (Gehring et al., 2011; Wolff et al., 2011). But RT-PCR and reporter analysis showed that the embryonic parent-of-origin-dependent expression pattern is not due to contamination. Therefore, we expect that a careful reexamination of the embryonic MEG and PEG candidates found in other studies could confirm additional genes regulated by genomic imprinting in the embryo, as is the case for the genes reported in this study.

Comparison of this study's 80 imprinted candidates with all embryonic MEG and PEG candidates identified in different studies reveals no genes to be common in all studies and only three overlapping genes between two studies (*AT2G01520*, *AT1G49450* and *AT1G57800*, Hsieh et al., 2011; Gehring et al., 2011; Nodine and Bartel, 2012). Similarly, the overlap between four recent studies (Hsieh et al., 2011; Gehring et al., 2011; Wolff et al., 2011; McKeown et al., 2011) identifying imprinted genes in the endosperm of *Arabidopsis* is minor, with only 20 genes of more than 300 being shared between two studies analyzing the same ecotypes and similar developmental stage (Pignatta and Gehring, 2012). The low number of common genes may be explained, at least to some extent, by the use of different accessions and the analysis of different developmental stages. Thus, some of the potential MEGs and PEGs, in both the embryo and endosperm, seem to be imprinted rather allele-specifically than locus-specifically and/or imprinting could be specific to a given developmental stage. In addition, different statistical procedures to select potentially imprinted genes may also contribute to the discrepancies (Deveale et al., 2012). Analyzing the datasets prior to stringent filtering increases the overlap (Wolff et al., 2011), as well as treating two different datasets with the same statistical pipeline (Gehring et al., 2011). Taken together, the recent, rapid development of high-throughput transcriptome sequencing allows large-scale identification of imprinted genes in different organisms and tissues. Yet, the obvious discrepancy between different datasets shows that in-depth analyses and substantial validation of imprinted candidate genes using alternative methods is necessary.

In conclusion, we show that at least 9 MEGs and one PEG are regulated by genomic imprinting in the embryo, and in addition, two MEGs display steady-state monoallelic expression in the embryo. Likely, the list of MEGs and PEGs in plant embryos will expand in the future by validating additional candidates or by high-

Table 2-3-2. Overview of the embryonic MEGs and the PEG.

GENE	GO-term		Monoallelic?	GUS in embryo	in gametes?	regulation
AT1G29660	lipid metabolism	MEG	YES	YES[3]	NO	PRC2
AT1G72260	defense (JA)	MEG	YES	YES[3]	YES	PRC2
AT2G47115	unknown	MEG	YES	YES[3]	not on array	?
AT5G62210	lipid metabolism	MEG	YES	YES[3]	NO	?
AT3G20520	lipid metabolism	MEG	YES	YES[3]	NO	?
AT2G17710	unknown	MEG	YES	YES[3]	NO	?
AT3G21500	terpenoid metabolism	MEG	YES	YES[4]	NO	?
AT2G01520	Meristem phase transition	MEG	YES[1]	ND[5]	NO	?
AT1G20680	unknown	MEG	YES[1]	ND[5]	NO	?
AT5G51950	Redox-function	MEG	YES[1]	ND[5]	NO	?
AT1G29050	unknown	MEG	YES[1]	ND[5]	NO	?
AT3G26790	embryogenesis	PEG	YES[2]	YES[6]	NO	PRC2

[1]: Monoallelic expression is only confirmed in 6x washed Col-0 x Ler control sample.
[2]: Monoallelic expression of *FUS3* (*AT3G26790*) only in early embryos (2-4 cell).
[3]: Reporter is expressed and imprinted in the embryo.
[4]: Reporter is weakly expressed in the embryo and imprinted expression was not assessed.
[5]: ND indicates that the reporter line is not available.
[6]: Reporter is expressed in the embryo but not imprinted.

throughput sequencing of additional embryonic samples.

PRC2 but not MET1 Is Involved in Maintaining Repression at the Silent Allele

The regulation of monoallelic and parent-of-origin-dependent gene expression largely depends on differential DNA methylation of the parental alleles in mammals and of some imprinted loci in the plant endosperm (Feil and Berger, 2007). We analyzed the effect of a paternal *met1-3* mutation, known to derepress silent paternal alleles of *FWA* and *FIS2* (Kinoshita et al., 2004; Jullien et al., 2006) in the endosperm, on imprinted gene expression in the embryo. We did not find any effect of paternal *met1-3* on the expression of the 11 MEGs, suggesting that MET1-mediated DNA-methylation in the CG context is not important for the regulation of embryonic MEGs (Figure 2-3-4, Figure S2-3-10). By contrast, *met1-3* abrogated paternal *FUS3* expression, indicating a role of *MET1* in activating paternal *FUS3* (Figure 2-3-4C). However, further studies are required to confirm this observation and to rule out that *FUS3* is not just below detection level in this sample as it is the case for two wild-type samples as well (Figure 2-3-1G).

The second, well-established imprinting regulator is PRC2, which mediates H3K27me3 (Raissig et al., 2011). In fact, we found that in embryos lacking maternal *FIE* activity, the usually silent paternal allele of two MEGs was derepressed (Figure 2-3-4A and 2-3-4B). Similarly, the imprinted genes *MEA* and *AtFH5* are biallelically expressed in seeds with a maternal mutation in a PRC2 subunit (Fitz Gerald et al., 2009; Jullien et al., 2006). In addition, the paternal *MEA* allele was shown to be derepressed by a paternal mutation in PRC2 components, which is not the case for the identified MEGs in the embryo (Jullien et al., 2006). However, this suggests that maternal PRC2 is involved in maintaining the silent state of paternal alleles of imprinted genes in both, the endosperm and the embryo. In addition, in embryos inheriting a paternal *fie* mutation, the expression pattern of the PEG was inverted: Instead of being paternally expressed, *FUS3* seems to be solely maternally expressed (Figure 2-3-4C). Our result suggests that paternal *FIE* is somehow required to activate the paternal *FUS3* allele, and that paternal *FUS3* is involved in the repression of the maternal *FUS3* allele just after fertilization. Such a negative feedback regulation of the gene product on its own imprinted expression has also been described for *MEA* (Baroux et al., 2006; Jullien et al., 2006). However, whereas paternal PRC2 seems to be involved in the activation

of the paternal *FUS3* allele, imprinted expression of *PHE1* requires maternal PRC2, which in this case is involved in the repression of the maternal allele of *PHE1* (Köhler et al., 2005). This indicates that the requirement of PRC2 for regulation of imprinted expression of PEGs differs between the two fertilization products.

We could not find any effect of mutations in *MET1* and *FIE* on 9 MEGs. Thus, other mechanisms must be involved in regulating the parent-of-origin-dependent gene expression in the embryo. As proposed for the endosperm (Wolff et al., 2011), asymmetric, non-CG DNA methylation could be involved in silencing the paternal alleles. However, asymmetric DNA methylation in the CHG context involves the SUVH4 methyltransferase *KYP* (Jackson et al., 2002) and all of the identified MEGs and PEGs are still imprinted in the mutant *kyp* x Ler embryonic library (Autran et al., 2011, Table S2-3-1). Thus, asymmetric DNA methylation in the CHG context seems unlikely to play a role in regulating the remaining 9 MEGs. Any other epigenetic mark could account for or add to the imprinted expression pattern in the embryo. Especially histone modifications might be of importance since they are more readily reversible than DNA methylation (Cedar and Bergman, 2009) and different modifications have been associated with imprinted genes in maize (Haun and Springer, 2008).

Resetting the Imprint

In mammals, imprinting marks are reset during germ line development. Very early during embryogenesis, as the germ line is set aside, epigenetic marks are erased and re-established according to the embryos' sex (Barlow, 2011; Feng et al., 2010). In plants, no germ line is set aside and the gametes develop very late from differentiated, sporophytic cells. In the endosperm, one-way control of imprinting is sufficient, since the endosperm does not contribute to the next generation. In contrast, imprints on embryonic MEGs and PEGs have to be reset. The only well-studied, imprinted gene in the plant embryo is the maize gene *Mee1* (Jahnke and Scholten, 2009). Interestingly, both alleles are fully methylated in the gametes and the maternal allele gets specifically demethylated in the zygote indicating an additional, yet undiscovered primary imprinting mark. During embryogenesis the maternal allele continuously regains methylation and consequently becomes silent. Thus, we do not know whether remethylation is cause or consequence or even involved in resetting the imprint and we cannot speculate about the time of imprint erasure. In rice, the monoallelic expression pattern of Os10g05750 is maintained throughout development in the endosperm, but in the embryo, Os10g05750 starts to be expressed biallelically from 8 DAP. This suggests erasure of the potential imprint during late embryogenesis (Luo et al., 2011). However, we do not know whether Os10g05750 is really imprinted in the embryo since only steady-state mRNA levels were analyzed and expression data of the gene in the gametes is not available.

In *Arabidopsis*, we found that all MEGs and PEGs show biallelic (n=10) or no (n=2) expression in the early seedling, thus, the imprint must be erased either late in embryogenesis or very early in vegetative development. When analyzing the expression pattern of the 11 embryonic MEGs and the PEG in the allele-specific dataset of torpedo-staged embryos (Gehring et al., 2011), we found that most genes are still expressed monoallelically or with a parental bias (n=8), although at very low levels. Thus, complete resetting of the imprint seems to occur after the torpedo stage for most of the MEGs.

However, how an imprint is erased during late embryogenesis and reset during gametogenesis is unknown. In the case of the two MEGs, where the repression of the maternal allele is maintained by PRC2, likely by the seed-specific MEA-FIS2 complex (Bemer and Grossniklaus, 2012), it is tempting to speculate that with decreasing expression of *MEA* during seed development (Baroux et al., 2006), activity of the MEA-FIS2 complex in the seed is decreasing and the H3K27me3 imprint might get lost by passive dilution during embryogenesis. However, future research will identify the primary imprinting mark(s) in the embryo, which will shed light on the yet unknown mechanism of resetting the imprint at embryonic MEGs and PEGs.

Biological Significance and Evolution of Genomic Imprinting in the Embryo

In placental mammals and in flowering plants, mutations in many imprinted genes cause growth defects in embryo and/or the nourishing tissue (ie. placenta or endosperm) in a parent-of-origin-specific manner (Grossniklaus et al., 1998; Ludwig et al., 1996; Lau et al., 1994; Ingouff et al., 2005; Kinoshita et al., 1999; Kiyosue et al., 1999; Luo et al., 2000; Tycko and Morison, 2002). Growth defects are consistent with a role of genomic imprinting in a parental conflict over resource allocation to the developing offspring (Haig and Westoby, 1989). Yet, when we analyzed available T-DNA lines disrupting embryonic MEGs, we could not find fertility, patterning or obvious growth phenotypes. This might be a result of gene redundancy and/or a subtle role of the genes during embryogenesis and seed development, which is not revealed in controlled and non-competitive laboratory conditions.

Notably, 5 of 11 MEGs have a role in metabolism, whereas 4 others are of unknown function (Table 2-3-2). In addition, all MEGs are expressed in the maternal seed coat and some MEG reporter lines and the PEG reporter line show a slightly biased expression towards the basal embryo and the suspensor (Figure 2-3-4, Figure S2-3-7). This suggests that embryonic MEGs might have a function at the interface between embryo and mother, possibly by linking seed coat metabolism and embryo metabolism and rendering the genes in the embryo under maternal control. This would be in line with the co-adaptation imprinting hypothesis: It predicts maternal expression of genes affecting mechanisms that are crucial at the maternal-offspring interface (Bateson, 1994; Wolf and Hager, 2006). In addition, the co-adaptation hypothesis predicts that the number of MEGs must be much higher than the number of PEGs, at least in species where the offspring develops within the mother. In fact, we find a large excess of MEGs (>90%), reminiscent of all other studies that analyzed parent-of-origin allelic expression in plant embryos. Both studies analyzing *Arabidopsis* embryos call 97% or 94% MEGs, respectively (Hsieh et al., 2011; Gehring et al., 2011). In rice only embryonic MEGs were called (Luo et al., 2011) and in maize embryos 76% of the imprinted candidates in the embryo show maternal expression (Waters et al., 2011). Also in the *Arabidopsis* and the rice endosperm more MEGs than PEGs were identified, but the fraction of embryonic MEGs is still higher than the fraction of endosperm-specific MEGs (Hsieh et al., 2011; Gehring et al., 2011; Wolff et al., 2011; McKeown et al., 2011). This suggests, that the co-adaptation hypothesis might be of importance for the evolution of genomic imprinting in the embryo, whereas the parental conflict might drive evolution in the nourishing tissue, the endosperm. Nevertheless, the evolution of genomic imprinting is likely due to a combination of parental conflict, mother-offspring co-adaptation and other factors depending on the locus and the tissue of expression.

In conclusion, we describe and confirm parent-of-origin-dependent, monoallelic expression in the *Arabidopsis* embryo. PRC2 is involved in the regulation of parent-of-origin allelic expression at some loci analyzed, but by far not all, suggesting additional, yet undiscovered regulators of genomic imprinting in the *Arabidopsis* embryo. Probably, the imprint is reset late in embryogenesis or early in vegetative development since all genes are either expressed from both parents or not at all in young seedlings. However, what the primary imprint is and when exactly and how it is reset, is currently unknown. Future research will likely confirm some of the embryonic MEGs and PEGs from other studies and help in the elucidation of the regulation and resetting of genomic imprinting in the embryo.

MATERIALS AND METHODS

Plant Material And Growth Conditions

Columbia-0 (Col-0) and Landsberg *erecta* (Ler) are the standard wild-type accession used in this study. We reciprocally crossed Col-0 and Ler to produce the hybrid embryonic samples and Col-0 was used for all *Agrobacterium*-mediated transformations in this study. The *fie/FIE* mutant (Col-0 background) used is SALK_042962 and the line has been described in detail in (Bouyer et al., 2011). The *met1-3/MET1* mutant (Col-0 background) used was first described in (Saze et al., 2003), and was only propagated heterozygously. It was assessed for full methylation at the 180 bp CEN-repeat by Southern blot analysis before crossing, indicating an unaltered epigenetic landscape and excluding uncontrollable, indirect effects (Wöhrmann et al., 2012). The *met1-3* genotyping assay is described in (Wöhrmann et al., 2012). All plants were grown in a greenhouse chamber with 16h light at ~20°C and 8h dark at ~18°C with an average of 60% humidity. For crosses, plants were emasculated and pollinated 2 days later.

Calling potentially imprinted genes in the embryo

The dataset from (Autran et al., 2011) was analyzed and we called all genes that had a q-value bigger than 0.8 (strong mono-parental bias, mi > 0.8 and pi > 0.8 for MEGs and PEGs, respectively) in all sequenced samples (2-4 cell Ler x Col-0, 2-4 cell *kyp/KYP* x Col-0, globular Ler x Col-0), in the 2-4cell samples only, or in the globular wild-type sample only. All filtered genes were then compared to the second replicate run and only kept if they still showed reads from one parent only (Autran et al., 2011). We also accepted genes that were sequenced in the globular sample of replicate 1 only (SOLiD 2009) and were not detected in the second replicate (SOLiD 2010). This procedure yielded 50 potential MEGs and 30 potential PEGs (Table S2-3-1). Expression levels (coverage by covered base Autran et al., 2011) and present/absent calls in the egg cell and sperm cell (Wuest et al., 2010; Borges et al., 2008) were used to prioritize the potential embryonic MEGs and PEGs. MEGs and PEGs being highly expressed and showing preferably absent calls in the gametes (egg cell or sperm cells) were selected for in-depth analysis (i.e. RT-PCR and Sanger sequencing).

Preparation of hybrid embryonic cDNA libraries

Different wild-type accessions and/or mutant lines were reciprocally crossed as indicated in the main text, the figures, and figure legends to produce hybrid F1 seeds. For the 1x-washed, wild-type embryonic samples we produced two independent biological replicates for each stage and direction of cross (i.e. 8 samples). The 2-4 cell embryos were isolated from seeds ~2.5 days after pollination (DAP), whereas the globular embryo stage was isolated from seeds ~4 DAP under our growth conditions. Embryo isolation was essentially performed as described in (Autran et al., 2011) with 5 additional washes after isolation for the extensively, 6x washed control samples (2-4 cell stage, reciprocally crossed). RNA was extracted using the Arcturus® PicoPure® RNA Isolation kit (Applied Biosystems) and the cDNA library ampflified using the Ovation® Pico WTA System (NuGEN) according to the manufacturer's protocol. As recommended by the Ovation® Pico WTA System (NuGEN) we purified the cDNA libraries with the QIAquick PCR Purification Kit (QIAGEN) according to NuGEN's protocol. We used the Agilent 2100 Bioanalyzer (Agilent Technologies) to control cDNA library quality and measured quantity using Nanodrop. In addition, we controlled library quality and absence of genomic DNA contamination by RT-PCR amplifying *ACTIN 11* (*ACT11*) and *WUSCHEL-RELATED HOMEOBOX 9* (*WOX9*), an embryo-specific gene (Wu et al., 2007). All primer sequences are specified in Table S2-3-5. In order to produce hybrid F1 seedling cDNA libraries and hybrid F1 seedling genomic DNA samples, we crossed Col-0 and Ler reciprocally, germinated the F1 hybrid seeds on plate and harvested them 8 days after sowing. Genomic DNA was extracted using the QiaQuick DNeasy kit (QIAGEN) and RNA was extracted using the NucleoSpin® RNA Plant Kit (Machery-Nagel). Reverse

transcription was essentially performed as previously published (Baroux et al., 2006).

RT-PCR and Sanger sequencing

RT-PCR was performed on diluted cDNA libraries (4ng/µl) by doing 28 to 34 cycles (94ºC for 15 sec, 58ºC for 20 sec, and 72ºC for 30 sec) followed by 72°C for 5min. We used Sigma Taq DNA Polymerase and PCR buffer from Sigma-Aldrich and a final concentration of 2mM $MgCl_2$, 0.2mM dNTPs and 0.2-0.4mM Primer. The resulting PCR product was analyzed on a standard DNA agarose gel and the remaining product was purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel). The purified PCR product was Sanger sequenced and the chromatograms analyzed at the site of the SNP between Ler and Col-0 to assess its parent-of-origin. All assays were tested for non-biased amplification of both accession sequence fragments using genomic DNA of F1 hybrid seedlings (Col-0 x Ler and Ler x Col-0). All sequences of the used primer are specified in Table S2-3-5.

Reporter lines: Cloning, Transformation, and Analysis

All GUS reporter lines were cloned using the pBGWFS7 vector (VIB, University of Gent), carrying a BASTA resistance gene (plant selection), a spectinomycin resistance gene (bacterial selection), and a Gateway-cloning cassette followed by eGFP and a *uidA* gene encoding ß-Glucuronidase (GUS) in frame. We amplified the upstream promoter region (from the previous gene until the start codon or a maximum of 2.5kb of promoter sequence) of seven MEGs and one PEG containing the attB recombination sites in a two-step PCR reaction: First, we used chimeric primers comprising template-specific sequences plus the first 12 bases of the attB1 or attB2 sequence at the 5'-end. PCR was performed with the Phusion High-Fidelity DNA Polymerase (Finnzymes) and buffer, 0.2mM dNTPs, 0.4µM Primers using the attB adapter program 1 (98°C for 60 sec; 5 cycles of 98°C for 10sec, 63°C for 20sec, 72°C for 60-180 sec; 30 cycles of 98°C for 10sec, 68°C for 20sec, 72°C for 60-180 sec; 72°C for 300sec). After analyzing the product on gel, we used 1 µl of the 50x diluted first PCR product as template for the second PCR using attB adapter primers (attB1-adaptor: 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCT-3' and attB2-adaptor: 5'-GGGGACCACTTTGTACAAGAAAGCTGGGT-3') to complete the attB sites. PCR was performed as above using the attB adapter program 2 (98°C for 60 sec; 5 cycles of 98°C for 10sec, 48°C for 20sec, 72°C for 60-180 sec; 15 cycles of 98°C for 10sec, 58°C for 20sec, 72°C for 60-180 sec; 72°C for 300sec). The resulting PCR product containing the promoter sequence and the complete attB recombination sites were PEG precipitated and the BP reaction (using pDONR221) and the LR reaction were performed according to the manufacturer's recommendations (Invitrogen). The resulting expression vectors were transformed into competent *Agrobacterium tumefaciens* (GV3101), which were used to transform Col-0 plants by floral dipping (Clough and Bent, 1998).

GUS reporter assays on isolated embryos

We first selected T1 lines strongly expressing the GUS reporter gene in the seed by staining young siliques overnight at 37°C after vacuum-infiltration (5-10 min) of the tissue in standard GUS stainining solution (2 mM 5-bromo,4-chloro,3-indolyl-D-glucuronide (Biosynth-AG), 10 mM EDTA, 0.1% Triton X-100, 2.0 mM potassium ferrocyanide, 2mM potassium ferricyanide, 50 mM phosphate buffer pH 7.2). The strong lines were selected, and reciprocally crossed with wild-type Col-0 plants. We then isolated embryos 2.5 DAP (2-4 cell stage) and 3.5 to 4 DAP (globular stage) in GUS staining solution (as above but with 0.5 mM Potassium ferro- and ferricyanide instead of 2.0 mM for improved GUS activity). We directly transferred the isolated embryos on a microscopy slide, added fresh GUS staining solution, covered the embryos with a coverslip and stained them without vacuum-infiltration for 4 d at 37°C in plastic boxes with high humidity to prevent drying of the samples. After 4 days we analyzed the isolated embryos for GUS reporter expression using bright-field microscopy (Leica DMR) to ensure

maximum sensitivity for GUS detection.

Mutant analysis

Available T-DNA insertion lines disrupting confirmed MEGs and PEGs (see Table S2-3-2) were ordered (2 lines/gene, if available). A mutant population (i.e. 24 individuals) was genotyped using primers flanking the insertion site (see Table S2-3-5, designed with the T-DNA primer design homepage http://signal.salk.edu/tdnaprimers.2.html) and the appropriate left boarder primer (for SALK lines: LBb1.3; for SAIL lines: Syg_LB1; for GABI lines: GBF_AC161_LB1; for FLAG lines: FL_LB4; for sequences see Table S2-3-5) using a standard PCR program (94ºC for 15 sec, 58ºC for 20 sec, and 72ºC for 75 sec, 36 cycles). Then, mature siliques of each genotyped individual were opened to analyze the seed set. In addition to that, we harvested siliques at different developmental stages of one (usually heterozygous) mutant individual, dissected the seeds in modified Hoyer's solution (70% w/v chloralhydrate, 4% w/v glycerol, 5% w/v gum arabic) and examined embryo patterning and development from the zygote to the torpedo stage using differential interference contrast (DIC) microscopy (Leica DMR).

ACKNOWLEDGEMENTS

We would like to thank Heike Lindner and Marian Bemer for critically reading the manuscript. We thank A. Schnittger and J. Paszkowski for seeds and O. Mittelsten Scheid for the 180-bp CEN repeat probe. MTR, CB and UG designed the experiments, MTR performed the experiments, MTR and UG analyzed the data and MTR, CB and UG wrote the manuscript. This work was supported by the University of Zürich and grants from the Swiss National Science Foundation and the European Research Council (to UG).

References

Autran, D., Baroux, C., Raissig, M.T., Lenormand, T., Wittig, M., Grob, S., Steimer, A., Barann, M., Klostermeier, U.C., Leblanc, O., Vielle-Calzada, J.-P., Rosenstiel, P., Grimanelli, D., and Grossniklaus, U. (2011). Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. Cell **145**: 707–719.

Barlow, D.P. (2011). Genomic imprinting: a mammalian epigenetic discovery model. Annu Rev Genet **45**: 379–403.

Baroux, C., Gagliardini, V., Page, D.R., and Grossniklaus, U. (2006). Dynamic regulatory interactions of *Polycomb* group genes: *MEDEA* autoregulation is required for imprinted gene expression in *Arabidopsis*. Genes Dev **20**: 1081–1086.

Bartolomei, M.S. and Ferguson-Smith, A.C. (2011). Mammalian genomic imprinting. Cold Spring Harb Perspect Biol **3**: 1-17.

Bateson, P. (1994). The dynamics of parent-offspring relationships in mammals. Trends Ecol Evol **9**: 399–403.

Bemer, M. and Grossniklaus, U. (2012). Dynamic regulation of *Polycomb* group activity during plant development. Curr Opin Plant Biol **15**: 523-9.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J.A., and Becker, J.D. (2008). Comparative transcriptomics of *Arabidopsis* sperm cells. Plant Phys **148**: 1168–1181.

Bouyer, D., Roudier, F., Heese, M., Andersen, E.D., Gey, D., Nowack, M.K., Goodrich, J., Renou, J.P., Grini, P.E., Colot, V., and Schnittger, A. (2011). *Polycomb* repressive complex 2 controls the embryo-to-seedling phase transition. PLoS Genet **7**: e1002014.

Cedar, H. and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet **10**: 295–304.

Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S., and Peacock, W.J. (1997). Fertilization-independent seed development in *Arabidopsis thaliana*. Proc Natl Acad Sci USA **94**: 4223–4228.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. Cell **110**: 33–42.

Clough, S.J. and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J **16**: 735–743.

Deveale, B., Van Der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA–Seq: A new perspective. PLoS Genet **8**: e1002600.

Feil, R. and Berger, F. (2007). Convergent evolution of genomic imprinting in plants and mammals. Trends Genet **23**: 192–199.

Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science **330**: 622–627.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet **12**: 565–575.

Fitz Gerald, J.N., Hui, P.S., and Berger, F. (2009). *Polycomb* group-dependent imprinting of the actin regulator *AtFH5* regulates morphogenesis in *Arabidopsis thaliana*. Development **136**: 3399–3404.

Frost, J.M. and Moore, G.E. (2010). The importance of imprinting in the human placenta. PLoS Genet **6**: e1001015.

Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes *MEDEA Polycomb* gene self-imprinting by allele-specific demethylation. Cell **124**: 495–506.

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. PLoS ONE **6**: e23687.

Grossniklaus, U., Vielle-Calzada, J.P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal control of embryogenesis by *MEDEA*, a *Polycomb* group gene in *Arabidopsis*. Science **280**: 446–450.

Gutiérrez-Marcos, J.F., Costa, L.M., Dal Prà, M., Scholten, S., Kranz, E., Perez, P., and Dickinson, H.G. (2006). Epigenetic asymmetry of imprinted genes in plant gametes. Nat Genet **38**: 876–878.

Haig, D. and Westoby, M. (1989). Parent-specific gene expression and the triploid endosperm. American Naturalist **134**: 147–155.

Haun, W.J., Laoueillé-Duprat, S., O'connell, M.J., Spillane, C., Grossniklaus, U., Phillips, A.R., Kaeppler, S.M., and Springer, N.M. (2007). Genomic imprinting, methylation and molecular evolution of maize *Enhancer of zeste* (*Mez*) homologs. Plant J **49**: 325–337.

Haun, W.J. and Springer, N.M. (2008). Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. Plant J **56**: 903–912.

Hermon, P., Srilunchang, K.O., Zou, J., Dresselhaus, T., and Danilevskaya, O.N. (2007). Activation of the imprinted *Polycomb* group *Fie1* gene in maize endosperm requires demethylation of the maternal allele. Plant Mol Biol **64**: 387–395.

Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., and Fischer, R.L. (2011). Regulation of imprinted gene expression in *Arabidopsis* endosperm. Proc Natl Acad Sci USA **108**: 1755–1762.

Ingouff, M., Haseloff, J., and Berger, F. (2005). *Polycomb* group genes control developmental timing of endosperm. Plant J **42**: 663–674.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature **416**: 556–60.

Jahnke, S. and Scholten, S. (2009). Epigenetic resetting of a gene imprinted in plant embryos. Curr Biol **19**: 1677–1681.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J **6**: 3901–7.

Jullien, P., Katz, A., Oliva, M., Ohad, N., and Berger, F. (2006). *Polycomb* group complexes self-regulate imprinting of the *Polycomb* group gene *MEDEA* in *Arabidopsis*. Curr Biol **16**: 486–492.

Jullien, P., Kinoshita, T., Ohad, N., and Berger, F. (2006). Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. Plant Cell **18**: 1360–1372.

Jullien, P.E. and Berger, F. (2009). Gamete-specific epigenetic mechanisms shape genomic imprinting. Curr Opin Plant Biol **12**: 637–642.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. Science **303**: 521–523.

Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Imprinting of the *MEDEA Polycomb* gene in the *Arabidopsis* endosperm. Plant Cell **11**: 1945–1952.

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Control of fertilization-independent endosperm development by the *MEDEA Polycomb* gene in *Arabidopsis*. Proc Natl Acad Sci USA **96**: 4186–4191.

Köhler, C., Page, D.R., Gagliardini, V., and Grossniklaus, U. (2005). The *Arabidopsis thaliana MEDEA Polycomb* group protein controls expression of *PHERES1* by parental imprinting. Nat Genet **37**: 28–30.

Lau, M.M., Stewart, C.E., Liu, Z., Bhatt, H., Rotwein, P., and Stewart, C.L. (1994). Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. Genes Dev **8**: 2953–2963.

Ludwig, T., Eggenschwiler, J., Fisher, P., D'Ercole, A.J., Davenport, M.L., and Efstratiadis, A. (1996). Mouse mutants lacking the type 2 IGF receptor (IGF2R) are rescued from perinatal lethality in *Igf2* and *Igf1r* null backgrounds. Dev Biol **177**: 517–535.

Luo, M., Bilodeau, P., Dennis, E.S., Peacock, W.J., and Chaudhury, A. (2000). Expression and parent-of-origin effects for *FIS2*, *MEA*, and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. Proc Natl Acad Sci USA **97**: 10637–10642.

Luo, M., Taylor, J., Spriggs, A., Zhang, H., Wu, X., Russell, S., Singh, M., and Koltunow, A. (2011). A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. PLoS Genet **7**: e1002125.

McKeown, P.C., Laouielle-Duprat, S., Prins, P., Wolff, P., Schmid, M.W., Donoghue, M.T., Fort, A., Duszynska, D., Comte, A., Lao, N.T., Wennblom, T.J., Smant, G., Köhler, C., Grossniklaus, U., and Spillane, C. (2011). Identification of imprinted genes subject to parent-of-origin specific expression in *Arabidopsis thaliana* seeds. BMC Plant Biol **11**: 113.

Nodine, M.D. and Bartel, D.P. (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature **482**: 94–97.

Pignatta, D. and Gehring, M. (2012). Imprinting meets genomics: new insights and new challenges. Curr Opin Plant Biol **15**: 530-5

Raissig, M.T., Baroux, C., and Grossniklaus, U. (2011). Regulation and flexibility of genomic imprinting during seed development. Plant Cell **23**: 16–26.

Raz, V., Bergervoet, J.H., and Koornneef, M. (2001). Sequential steps for developmental arrest in *Arabidopsis* seeds. Development **128**: 243–252.

Reik, W. and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat Rev Genet **2**: 21–32.

Saze, H., Mittelsten Scheid, O., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat Genet **34**: 65–69.

Spillane, C., Schmid, K.J., Laoueillé-Duprat, S., Pien, S., Escobar-Restrepo, J.M., Baroux, C., Gagliardini, V., Page, D.R., Wolfe, K.H., and Grossniklaus, U. (2007). Positive darwinian selection at the imprinted *MEDEA* locus in plants. Nature **448**: 349–352.

Tiwari, S., Schulz, R., Ikeda, Y., Dytham, L., Bravo, J., Mathers, L., Spielman, M., Guzmán, P., Oakey, R.J., Kinoshita, T., and Scott, R.J. (2008). *MATERNALLY EXPRESSED PAB C-TERMINAL*, a novel imprinted gene in *Arabidopsis*, encodes the conserved C-terminal domain of polyadenylate binding proteins. Plant Cell **20**: 2387–2398.

Tycko, B. and Morison, I.M. (2002). Physiological functions of imprinted genes. J Cell Physiol **192**: 245–258.

Waters, A.J., Makarevitch, I., Eichten, S.R., Swanson-Wagner, R.A., Yeh, C.T., Xu, W., Schnable, P.S., Vaughn, M.W., Gehring, M., and Springer, N.M. (2011). Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. Plant Cell **23**: 4221–4233.

Wolf, J.B. and Hager, R. (2006). A maternal-offspring coadaptation theory for the evolution of genomic imprinting. PLoS Biol **4**: e380.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C. (2011). High-resolutionanalysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. PLoS Genet **7**: e1002126.

Wu, X., Chory, J., and Weigel, D. (2007). Combinations of WOX activities regulate tissue proliferation during *Arabidopsis* embryonic development. Dev Biol **309**: 306–316.

Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenführer, J., Von Mering, C., and Grossniklaus, U. (2010). *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. Curr Biol **20**: 506–512.

Wöhrmann, H.J., Gagliardini, V., Raissig, M.T., Wehrle, W., Arand, J., Schmidt, A., Tierling, S., Page, D.R., Schöb, H., Walter, J., and Grossniklaus, U. (2012). Identification of a DNA methylation-independent imprinting control region at the *Arabidopsis MEDEA* locus. Genes Dev **26**: 1837–1850.

Xiang, D., Venglat, P., Tibiche, C., Yang, H., Risseeuw, E., Cao, Y., Babic, V., Cloutier, M., Keller, W., Wang, E., Selvaraj, G., and Datla, R. (2011). Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*. Plant Phys **156**: 346–356.

Zhang, M., Zhao, H., Xie, S., Chen, J., Xu, Y., Wang, K., Guan, H., Hu, X., Jiao, Y., Song, W., and Lai, J. (2011). Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. Proc Natl Acad Sci USA **108**: 20042-7.

DISCUSSION

GENERAL DISCUSSION AND FUTURE PERSPECTIVES

NOTE

All findings of this thesis are already discussed in detail at the end of each chapter. The last chapter “General Discussion and Future Perspectives” summarizes again shortly the main findings and its implications, and concludes by discussing future directions and fundamental open questions regarding the identification of novel imprinted plant genes, the tissue-specific occurrence of plant imprinting, and the regulation of imprinted expression in plants, which is far away from being solved. In addition and foremost, technical challenges and desirable methodological advancements to solve unanswered issues due to technical limitations are discussed.

Imprinting control regions (ICRs) - DNA methylation and other primary imprinting marks

Since ICRs had not been described in plant genomes, the identification of the short, unmethylated *MEA*-ICR was a major breakthrough in plant imprinting research. However, the first plant ICR, the *MEA*-ICR, possesses very distinct features compared to mammalian ICRs. The *MEA*-ICR is short, 200bp only, whereas mammalian ICRs span an average of 3kbs, and the *MEA*-ICR controls a singleton rather than a whole gene cluster that can cover up to 850 kb in the case of the *Kcnq1* cluster (Barlow, 2011; Wöhrmann et al., 2012). Furthermore, the *MEA*-ICR region is not targeted by DNA methylation, neither in the sperm cells, the central cell, nor in young embryos (Wöhrmann et al., 2012). In mammals, differential DNA methylation at ICRs regulates genomic imprinting either through an insulator function by DNA methylation-dependent binding of *trans*-acting factors (Bell & Felsenfeld, 2000; Hark et al., 2000; Szabó et al., 2000) or, if unmethylated, by acting as promoter of a ncRNA, which itself represses genes in *cis* (Lyle et al., 2000; Mancini-Dinardo et al., 2003; Wutz et al., 1997). Therefore, the only identified plant ICR seems to regulate genomic imprinting differently than those in mammals, and independently of DNA methylation. A two-level control mechanism was proposed for *MEA* imprinting regulation, where DNA methylation, and its effectors in plants, DME and MET1, control higher-order chromatin structure to allow or restrict access of additional *trans*-acting and *MEA*-ICR-binding imprinting regulators (Wöhrmann et al., 2012).

Obviously, DNA methylation is not the primary imprinting mark regulating imprinted *MEA* expression. Generally, not much is known about hierarchies of imprinting marks. For instance, it is not known how the *de novo* DNA methyltransferase DNMT3A and its cofactor DNMT3L recognize ICRs in the germ line to set up sex-specific germline methylation patterns in mammals. Although specific spacing of CpG islands in the ICR, histone modifications, or parent-specific transcription over ICRs seem to be involved (Chotalia et al., 2009; Ciccone et al., 2009; Glass et al., 2009; Ooi et al., 2007), the matter remains unresolved. In addition, hypomethylation of the *Snrpn* ICR caused by a mutation of the transcriptional repressor *Zfp57* in the maternal germline could be rescued by zygotic expression of ZFP57 and proper DNA methylation was reestablished during embryogenesis suggesting a distinct primary imprinting mark at the *Snrpn* ICR (Li et al., 2008). In comparison, the two maize imprinted genes *ZmFie1* and *ZmFie2* are regulated by differential DNA methylation, but only for one gene it is set up in the gametes already (Gutiérrez-Marcos et al., 2006). Furthermore, *FWA,* although being the prime example of one-way imprinting control by DNA methylation in plants (Kinoshita et al., 2004; Kinoshita et al., 2007), is not differentially methylated in the gametes (Wöhrmann et al., 2012), indicating that the differentially methylated regions (DMRs) are set up after fertilization. Taken together, primary or additional imprinting marks other than DNA methylation must be involved in regulating imprinted expression in both plants and animals, and will hopefully be identified by future research.

(Imperfect) Attempts to identify novel *trans*-acting factors at the *MEA*-ICR

In an attempt to identify the *trans*-acting factor(s) binding and/or regulating the *MEA*-ICR, we performed a biochemical and a forward genetic screen. An electrophoretic mobility shift assay (EMSA) was used to screen a broccoli nuclear protein extract for effectors binding to parts of the radiolabeled *MEA*-ICR, but failed due to plethora of uncontrollable aspects. The forward genetic screen was based on non-activation or ectopic activation of the *MEA*-ICR reporter gene *250pMEA::GUS* after EMS mutagenesis. Genetically, we could identify two mutant candidates that are involved in the activation of the reporter gene and are required for seed development. Unfortunately, the genetic background of the *250pMEA::GUS* line was very heterogeneous and differed from one individual to another even in the parental, unmutagenized generation. A prior gamma-ray irradiation and insufficient rounds of backcrosses contributed to the genetic heterogeneity, and, therefore, polymorphisms still segregated within the parental population. We wrongly assumed that we were working with an inbred paternal

line and realized it only by resequencing the parental line in order to create an *in-silico*-corrected, high-quality reference genome to map the resequenced mutant candidate reads for SRM. A low quality reference genome resulting in bad mapping quality concomitant with a very heterogeneous genetic background made it impossible to pinpoint the causative mutation. Therefore, we decided to outcross the mutation to inbred Col-0 plants to create near isogenic lines (NILs) and retry the mapping of the causative mutation. In general, we recommend performing forward genetic screens in accessions with a published high-quality reference genome, fully annotated and properly aligned, preventing any bioinformatic obstacles during the mapping procedure.

SNP ratio mapping: mapping heterozygous mutations by next-generation sequencing

We developed SNP ratio mapping (SRM) to map gametophytic-lethal or embryo-lethal mutations, which were isolated from two different forward genetic screens - one aiming at uncovering regulators of imprinted *MEA* expression (Chapter 1.2) and another to find additional factors involved in the pollen tube reception pathway (Chapter 1.3; Lindner et al. 2012) - in our laboratory. SRM allows mapping of homozygous-lethal mutations or of mutations in complex genetic and transgenic backgrounds. Lindner and colleagues successfully mapped three different mutations affecting the pollen tube reception pathway (Lindner et al., 2012; Lindner, Grossniklaus, unpublished). The need of few individuals and only two rounds of backcrosses after mutagenesis makes SRM a useful method for many genetic organisms, especially those with a long generation time like *Arabidopsis*. In addition, any kind of mutation can be mapped, in particular those, which show low transmission rates, are lethal, or cause hard-to-score phenotypes. Unlike other methods that either require many rounds of backcrosses (Zuryn et al., 2010) or homozygous viable mutations (Austin et al., 2011; Schneeberger et al., 2009) we present a method that allows the mapping of a mutation in the heterozygous state after only two rounds of backcrosses.

Efficient embryo isolation enabled the study of parental effects in plant embryos

The development of a method to rapidly and efficiently isolate young *Arabidopsis* embryos was essential to study parental effects during plant embryogenesis. While manual dissection is applicable for later-staged embryos (starting from late globular stage), dissection of young *Arabidopsis* embryos is hindered by their small size and deep embedding within seed coat and endosperm. We successfully performed RNA sequencing (RNA-seq) and allele-specific transcript analysis using RT-PCR and Sanger sequencing on amplified embryonic cDNA samples, assessed reporter gene expression on isolated embryos, and evaluated cytological aspects of plant embryos by fluorescence *in situ* hybridization (FISH). Obviously, we are far away from collecting a sufficient number of embryonic cells to apply genomics approaches like chromatin-immunoprecipitation followed by next-generation sequencing (ChIP-Seq), because chromatin, unlike RNA or DNA, cannot be amplified. ChIP-Seq approaches are essential to profile the epigenetic landscape genome-wide or at specific loci such as imprinted genes. Description of the epigenetic modifications at and around imprinted genes might add to our understanding of the regulatory processes and mechanisms involved that regulate imprinted expression (see below). However, recent publications performing ChIP-seq on plant tissue use 1 and 2 g of total plant tissue to perform ChIP (Zhong et al., 2012; Immink et al., 2012), as it is also suggested by a recent plant ChIP protocol (Villar and Köhler, 2010). Assuming that a cell consists only of water (density = 10^3 kg/m^3) and an embryonic cell is approximately 10^{-15} m^3 in volume (see Figure 2-1-4; Chapter 2.1) then one embryonic cell has a mass of 1 ng. Thus, 1 g of tissue corresponds to 1 billion embryonic cells, an impossible number to collect by hand. Successful isolation would require a more automated approach, such as fluorescence-assisted cell sorting (FACS) of embryonic cells expressing a fluorescent reporter gene. Surely, plant research would highly benefit of more sophisticated high-throughput systems to isolate specific plant cell types (see below).

Maternal dominance or equal parental policy during embryo development in *Arabidopsis*?

In 2011, we presented a genome-wide view of the parental contributions to early plant embryogenesis and described a gradually decreasing maternal dominance (Autran et al., 2011). However, we could not distinguish, whether the maternal dominance arises due to differential, *de novo* gene expression in the embryo or is caused by deposition of maternal transcripts in the cytoplasm of the egg cell, as it is the case for many animal systems (Baroux et al., 2009). In addition, genetic studies suggested a role for two maternal, epigenetic pathways in regulating paternal contribution. The 24nt siRNA pathway, or the RNA-dependent DNA methylation (RdDM) pathway, seems to be involved in repressing early paternal activity, and, to our knowledge, we were the first lab to report that this pathway not only targets repetitive and transposable elements but also protein-coding genes in this specific developmental context in plants (Autran et al., 2011; Matzke et al., 2009). Maternal siRNAs might have a role in heterosis and inter-specific hybridization (Chen, 2010; Martienssen, 2010), and it remains to be discovered whether RdDM mutants are sensitized to hybridization (Autran et al., 2011). Furthermore, histone turnover regulated by the maternal CAF1 complex, a chromatin assembly complex, aids in activating the paternal genome (Autran et al., 2011). Interestingly, early maternal dominance and gradual zygotic genome activation explains the developmental stage-dependent occurrence of embryo-lethal mutant phenotypes, which at early stages seem independent of the paternal genotype. Both mutations, *gnom/emb30* and *vacuoleless1* (*vcl1*), are classified as recessive and embryo-lethal (Mayer et al., 1993; Rojo et al., 2001). Surprisingly, the incidence of mutant phenotypes at early embryonic stages is not or only partially rescued by inheriting a paternal wild-type allele, suggesting that those genes are not or only weakly expressed paternally at early developmental stages, therefore not complementing the phenotype (Baroux et al., 2009).

In contrast, a recent paper describes equal parental contributions starting already from the 1-cell embryo (Nodine and Bartel, 2012). The authors attribute our result to substantial contamination with maternal, sporophytic tissue, which obviously could explain the strong maternal bias we observed (Autran et al., 2011). Yet, they fail to explain how we can observe the substantial increase in paternally contributed transcripts in embryonic samples produced from mutant *kryptonite* (*kyp*) mothers. The SUVH4 histone methyltransferase KYP (Jackson et al., 2002) was found to be involved in repressing paternal gene expression during early embryonic stages (Autran et al., 2011). Thus, we found double the amount of paternal information in both *kyp* x Col-0 embryo samples analyzed. Since we produced the mutant replicates exactly in the same way as the wild-type samples, it is difficult to explain why we have consistently less "contamination" in the mutant samples.

In addition, interesting biological aspects might underlie the differences found in the two studies. First, Nodine and Bartel (2012) used the Cape Verde Islands (Cvi) accession, being very different from all other accessions (Nordborg et al., 2005) and showing distinct epigenetic peculiarities (Gazzani et al., 2003; Riddle & Richards, 2002; Saze & Kakutani, 2007; Tessadori et al., 2009). Second, Nodine and Bartel (2012) analyzed polyadenylated mRNAs, whereas we profiled mRNAs irrespective of their polyadenylation status. In animals, cytoplasmic polyadenylation of maternal transcripts regulates its translation during early development (Benoit et al., 2008; Galili et al., 1988; Lasko, 2009) and, therefore, distinct amplification methods might add to the observed differences.

Nevertheless, resequencing of hybrid embryonic libraries produced after crosses with different ecotypes, including Col-0, Cvi, and Ler, and amplified specifically for polyadenylated transcripts or not, will shed light on the (yet) unresolved issue of parental policy during early plant embryogenesis.

Genomic imprinting in the *Arabidopsis* embryo

Even though the two above-mentioned publications do not agree on the overall parental contribution to early embryogenesis in *Arabidopsis*, both independently report parent-of-origin-dependent allelic gene expression.

Table D1. All identified candidate imprinted genes in plant embryos.

ARABIDOPSIS THALIANA

locus	parent-of-origin	stage	confirmed?
AT3G03750	paternal[1]	6 DAP	NO
AT5G38160	maternal[1]	6 DAP	NO
AT5G42235	maternal[1]	6 DAP	NO
AT1G20500	maternal[1]	6 DAP	NO
AT3G08900	maternal[1]	6 DAP	NO
AT1G62340	maternal[1]	6 DAP	NO
AT2G27380	maternal[1]	6 DAP	NO
AT5G50090	maternal[1]	6 DAP	NO
AT4G26420	maternal[1]	6 DAP	NO
AT1G05620	maternal[1]	6 DAP	NO
AT5G40360	maternal[1]	6 DAP	NO
AT1G71250	maternal[1]	6 DAP	NO
AT3G12590	maternal[1]	6 DAP	NO
AT3G28750	maternal[1]	6 DAP	NO
AT5G42930	maternal[1]	6 DAP	NO
AT1G72940	maternal[1]	6 DAP	NO
AT1G69520	maternal[1]	6 DAP	NO
AT1G46554	maternal[1]	6 DAP	NO
AT1G57800	paternal[2]	7-8 DAP	NO
AT1G12010	maternal[2]	7-8 DAP	NO
AT1G19640	maternal[2]	7-8 DAP	NO
AT1G23200	maternal[2]	7-8 DAP	NO
AT1G28590	maternal[2]	7-8 DAP	NO
AT1G49450	maternal[2]	7-8 DAP	NO
AT1G53282	maternal[2]	7-8 DAP	NO
AT1G62000	maternal[2]	7-8 DAP	NO
AT1G62060	maternal[2]	7-8 DAP	NO
AT1G62070	maternal[2]	7-8 DAP	NO
AT1G62080	maternal[2]	7-8 DAP	NO
AT1G69870	maternal[2]	7-8 DAP	NO
AT1G70140	maternal[2]	7-8 DAP	NO
AT1G70830	maternal[2]	7-8 DAP	NO
AT1G71250	maternal[2]	7-8 DAP	NO
AT2G21420	maternal[2]	7-8 DAP	NO
AT2G23580	maternal[2]	7-8 DAP	NO
AT2G25450	maternal[2]	7-8 DAP	NO
AT2G25540	maternal[2]	7-8 DAP	NO
AT2G27310	maternal[2]	7-8 DAP	NO
AT2G27390	maternal[2]	7-8 DAP	NO
AT2G28650	maternal[2]	7-8 DAP	NO
AT2G36760	maternal[2]	7-8 DAP	NO
AT3G22910	maternal[2]	7-8 DAP	NO
AT4G00220	maternal[2]	7-8 DAP	NO
AT4G09820	maternal[2]	7-8 DAP	NO
AT4G12290	maternal[2]	7-8 DAP	NO
AT4G12960	maternal[2]	7-8 DAP	NO
AT4G29230	maternal[2]	7-8 DAP	NO
AT4G37520	maternal[2]	7-8 DAP	NO
AT5G42235	maternal[2]	7-8 DAP	NO
AT5G42930	maternal[2]	7-8 DAP	NO
AT5G47150	maternal[2]	7-8 DAP	NO
AT5G47330	maternal[2]	7-8 DAP	NO
AT5G48100	maternal[2]	7-8 DAP	NO
AT5G53870	maternal[2]	7-8 DAP	NO
AT5G59310	maternal[2]	7-8 DAP	NO
AT5G63800	maternal[2]	7-8 DAP	NO

ORYZA SATIVA

locus	parent-of-origin	stage	confirmed?
Os02g06830.1	maternal[3]	6 DAP	NO
Os05g06480.1	maternal[3]	6 DAP	NO
Os08g08960.1	maternal[3]	6 DAP	NO
Os03g07660.1	maternal[3]	6 DAP	NO
Os06g23530.1	maternal[3]	6 DAP	NO
Os08g04150.1	maternal[3]	6 DAP	NO
Os10g05750.1	**maternal[3]**	**6 DAP**	**YES**

ZEA MAYS

locus	parent-of-origin	stage	confirmed?
GRMZM2G003909	maternal[4]	14 DAP	NO
GRMZM2G009465	maternal[4]	14 DAP	NO
GRMZM2G014119	maternal[4]	14 DAP	NO
GRMZM2G016145	maternal[4]	14 DAP	NO
GRMZM2G044440	maternal[4]	14 DAP	NO
GRMZM2G062650	maternal[4]	14 DAP	NO
GRMZM2G063498	maternal[4]	14 DAP	NO
GRMZM2G073700	maternal[4]	14 DAP	NO
GRMZM2G099960	maternal[4]	14 DAP	NO
GRMZM2G103247	maternal[4]	14 DAP	NO
GRMZM2G114356	maternal[4]	14 DAP	NO
GRMZM2G118205	maternal[4]	14 DAP	NO
GRMZM2G130580	maternal[4]	14 DAP	NO
GRMZM2G137502	maternal[4]	14 DAP	NO
GRMZM2G141382	maternal[4]	14 DAP	NO
GRMZM2G150134	maternal[4]	14 DAP	NO
GRMZM2G150680	maternal[4]	14 DAP	NO
GRMZM2G159636	maternal[4]	14 DAP	NO
GRMZM2G160687	maternal[4]	14 DAP	NO
GRMZM2G167630	maternal[4]	14 DAP	NO
GRMZM2G169695	maternal[4]	14 DAP	NO
GRMZM2G170099	maternal[4]	14 DAP	NO
GRMZM2G343972	maternal[4]	14 DAP	NO
GRMZM2G351505	maternal[4]	14 DAP	NO
GRMZM2G354579	maternal[4]	14 DAP	NO
GRMZM2G370991	maternal[4]	14 DAP	NO
GRMZM2G394203	maternal[4]	14 DAP	NO
AC191534.3_FG003	paternal[4]	14 DAP	NO
AC217300.3_FG004	paternal[4]	14 DAP	NO
GRMZM2G006732	paternal[4]	14 DAP	NO
GRMZM2G028366	paternal[4]	14 DAP	NO
GRMZM2G047104	paternal[4]	14 DAP	NO
GRMZM2G093947	paternal[4]	14 DAP	NO
GRMZM2G127160	paternal[4]	14 DAP	NO
GRMZM2G324131	paternal[4]	14 DAP	NO
GRMZM2G449489	paternal[4]	14 DAP	NO
GRMZM5G830365	paternal[4]	14 DAP	NO

[1]: Gehring et al., 2011
[2]: Hsieh et al., 2011
[3]: Luo et al., 2011
[4]: Waters et al., 2011

We analyzed the allele-specific transcriptome dataset of Autran and colleagues (2011) in detail and filtered all transcripts that show read counts from one parent but not from the other. We identified 50 MEG and 30 PEG candidates and studied 18 MEGs and 6 PEGs showing relatively high expression values and preferably absence

of expression in the gametes in detail. We confirmed 11 MEGs and 1 PEG by alternative methods, namely allele-specific transcript analyses and reporter gene analyses. Genetic studies showed that PRC2 is involved in regulating imprinted expression of 2 MEGs, *AT1G29660* and *AT1G72260*, and the PEG *FUS3* in the embryo. Mutant analysis of MEGs and the PEG revealed no obvious phenotype in terms of fertility and/or embryo patterning. However, five of eleven MEGs have a (predicted) function in metabolism, whereas another 4 genes are of unknown function. In addition, the MEGs are expressed in the whole seed, including the seed coat. This suggests that genomic imprinting in the embryo could have evolved to link maternal and embryonic metabolism and rendered the process even in the embryo under maternal control. This is in accordance with the co-adaptation theory (Wolf and Hager, 2006), which predicts that traits important at the mother-offspring interface have evolved to be under maternal control. Importantly, some MEG reporter lines have a slight bias of expression in the suspensor, the file of cells linking the embryo proper to the maternal sporophytic seed coat. Lastly, we have shown that the imprint must be erased late in embryo or early in vegetative development, since all embryonic imprinted genes are expressed from both parents in young seedlings. In conclusion, genomic imprinting in *Arabidopsis* exists in the embryo, albeit just transiently, and, thus, the current dogma that one-way imprinting in the endosperm is the only form of genomic imprinting in *Arabidopsis* is strongly put into question. Yet, what the primary imprinting mark is, how and where it is established, and how and when exactly it is erased, remains unknown.

Importantly, only 50% (12 of 24) of our "best" candidates turned out to be imprinted in the embryo, which indicates, that at least 50% of our totally 80 candidate MEGs and PEGs are false positives. Likely, the fraction of false positives is even bigger, since we only chose highly expressed genes that were covered by sufficient reads for an in-depth analysis. As described in the introduction, allele-specific genomics approaches have to be considered with a certain caution, since the sequencing depth and the statistical procedures to call potentially imprinted genes differ a lot between studies, and the commonly identified imprinted candidate genes in similar studies represent a minute fraction. For example, analysis of an allele-specific transcriptome of *Arabidopsis* seeds analyzing the same accessions and the same developmental stage only identified 20 common genes in 126 and 208 candidate imprinted genes, respectively (Pignatta and Gehring, 2012). In addition, a recent study showed that empirical estimation of false discovery rates (FDR) might be indispensable for next-generation genomics approaches to determine the allele-specific origin of transcripts (Deveale et al., 2012). Therefore, MEGs and PEGs identified by allele-specific genomics approaches have to be considered with caution before imprinted candidates are not substantially validated by alternative methods, or before a given imprinted candidate gene is not called several times in independent studies.

Tissue-specific imprinting in plants?

The identification of genes showing tissue-specific, imprinted expression in the embryo challenges the assumption that imprinted genes are preferentially expressed in the endosperm. Recent studies filtered an allele-specific seed transcriptome for genes specifically expressed in the endosperm (Wolff et al., 2011; Hsieh et al., 2011). Obviously, this is necessary when analyzing a mixture of different tissues, yet, imprinted genes that are expressed in the whole seed but might show lineage-specific imprinting, as it is the case for the embryonic MEGs we identified, would be missed. In addition, some of those studies also analyzed an embryonic allele-specific transcriptome and without exception found imprinted expression in the embryo (Table D1; Gehring et al., 2011; Hsieh et al., 2011; Luo et al., 2011; Waters et al., 2011). But except for the embryonic imprinted genes in rice, where one of 8 was shown to have a steady-state monoallelic expression profile by alternative methods (Table D1; Luo et al., 2011), all other studies discarded this finding with the argument that the genes are also expressed, and some are even imprinted, in the endosperm and, therefore, are considered to be contamination (Hsieh et al., 2011, Gehring et al., 2011, Waters et al., 2011). At least, Waters and colleagues (2011) present other potential

explanations and admit that separating endosperm and embryo 14 DAP in maize is easy and should not result in substantial endosperm contamination in the embryo sample. They speculated that the genes could be imprinted in both the embryo and the endosperm, as it is the case for the maize *Mee1* gene (Jahnke and Scholten, 2009), or transcripts might be remnants of very stable RNAs stored in the gametes, or might even be transported from endosperm to the embryo (Waters et al., 2011). However, in mammals lineage-specific and even dynamic regulation of imprinted genes is described. For example, the distal genes in the *Kcnq1* cluster are maternally expressed in the placenta, but biallelically contributed in the embryo and the adult mouse (Hudson et al., 2010). In addition, the mammalian gene *Dlk1* is paternally expressed during embryogenesis, but biallelically contributed in the neurogenic niche concomitant with lineage-specific methylation of the maternal ICR (Ferrón et al., 2011). Therefore, tissue-specific imprinting should be considered more in plants, since we showed that genes that are expressed throughout the seed exhibit imprinted expression in the embryo. Even if a gene is expressed in many different tissues, and is not restricted to embryo and/or endosperm, it could be imprinted in one specific lineage but show biallelic expression elsewhere.

Parent-of-origin-dependent allelic expression is largely absent in plant seedlings

In *Arabidopsis*, maize and rice, hybrid seedlings were analyzed for parental bias and the imprinted expression of genes during vegetative seedling development. Zhang and Borevitz (2009) used a high-density SNP-tilling array, covering 250'000 SNPs of 20 different ecotypes, to quantitatively assess over 12'000 genes for parental bias. At the analyzed threshold they did not identify any genes regulated by genomic imprinting in the seedling, suggesting that genomic imprinting is restricted to sexual development in *Arabidopsis* (Zhang and Borevitz, 2009). However, the authors analyzed only about half of all *Arabidopsis* genes and enriched for polyadenylated transcripts. In maize, Springer and Stupar (2007) analyzed hybrid seedlings, hybrid ears (female reproductive organ) and hybrid embryos 19 DAP. Whereas hardly any parental bias was identified in ears and seedlings, 20% of the analyzed genes in the embryo showed a bias towards the maternal allele, although the authors rather attributed the effect to incomplete maternal-to-zygotic transition (MZT) than to imprinting (Springer and Stupar, 2007). Lastly, He and colleagues (2010) performed RNA-seq and ChIP-seq on rice hybrid seedlings. They describe allelic bias in gene expression not in a parent-of-origin but rather in an ecotype-of-origin-dependent manner that is correlated to gene expression and epigenetic differences in the parental lines (He et al., 2010). These studies are in line with the finding that the identified embryonic MEGs and the PEG are biallelically contributed in the seedling. Taken together, imprinted gene expression seems to be largely absent from vegetative plant tissue, again suggesting that all embryonic imprints are erased likely during late embryogenesis and before vegetative development.

Novel regulatory players and the hierarchy of epigenetic modifications

Most imprinted gene clusters in mammals express a long ncRNA and most of them are directly involved the repression of genes in *cis* by recruiting a repressive epigenetic machinery (Santoro and Barlow, 2011). In both, maize and rice, recent studies described a number of imprinted ncRNA in the seed (Zhang et al., 2011; Luo et al., 2011), and four maternally expressed ncRNAs in maize are even transcribed from within PEGs, indicating a similar role for plant ncRNAs in the regulation of genomic imprinting as in mammals (Zhang et al., 2011). Yet, whether those ncRNAs are able to recruit the silencing machinery to genes in *cis* is unknown. However, a different long ncRNA, COLDAIR, regulates flowering time upon a period of cold by recruiting PRC2 to silence a floral repressor (Heo and Sung, 2010). In conclusion, the presence of imprinted ncRNAs in plants and the ability of non-imprinted ncRNAs to recruit a repressive epigenetic machinery to silence a gene in *cis* suggest a similar imprinting control mechanism as in mammals and should be in the focus of future research.

Furthermore, epigenetic profiling of chromatin at and around imprinted loci was only performed in maize but

not in *Arabidopsis* or rice. Haun and Springer (2008) analyzed three maternally expressed genes, *Mez1*, *ZmFie1* and *Nrp1*, and found enrichment of di- and trimethylation of H3K27 at paternal alleles, and enrichment of acetylated H3 and H4 and dimethylation at H3K4 at maternal alleles. Importantly, those marks are only associated with the alleles in the tissue that exhibits imprinting (the endosperm) but not in the seedling (Haun and Springer, 2008). Surprisingly, di- and trimethylation of H3K9, which is involved in silencing paternal alleles in mammals (Pandey et al., 2008; Redrup et al., 2009; Terranova et al., 2008) was not associated with any of the alleles. Thus, profiling the epigenetic status at imprinted loci in different developmental contexts might shed light (i) on involved regulatory machineries, (ii) on the hierarchical and time-dependent order of chromatin marks, and (iii) on differences and commonalities in regulating genomic imprinting in organisms as evolutionary distinct as plants and mammals.

Accession-specific, species-specific and kingdom-specific imprints

The first imprinted gene discovered, the maize *R1*-gene (Kermicle, 1970), as well as other maize genes are only imprinted in some accessions but not in others (Kermicle and Alleman, 1990; Messing and Grossniklaus, 1999), therefore exhibiting rather allele-specific than locus-specific imprinting. A recent study in *Arabidopsis* also describes several genes to be imprinted in one accession but not the other (Wolff et al., 2011). The extent of natural variation and the progress in sequencing the different accessions of *Arabidopsis thaliana* offers the possibility to investigate accession-dependent imprinting in this species. This might allow to specifically compare the epigenetic landscape around accession-dependent imprinted loci in an accession with imprinted and an accession with biallelic gene expression of a given locus, and could pinpoint causal differences in the (epigenetic) regulation leading to allele-specific expression. Furthermore, genome-wide association studies (Atwell et al., 2010) might uncover epigenetic regulators by correlating SNPs to the expression status - imprinted or not - of a given imprinted gene. Therefore, natural variation in plants and especially in *Arabidopsis* with many accessions' genomes being resequenced at the moment (www.1001genomes.org) offers possibilities to study imprinting in an evolutionary perspective. Interestingly, recent studies that profiled allele-specific seed transcriptomes to identify imprinted genes in maize, rice, and *Arabidopsis* (Gehring et al., 2011; McKeown et al., 2011; Hsieh et al., 2011; Wolff et al., 2011; Zhang et al., 2011; Waters et al., 2011; Luo et al., 2011), albeit finding a small overlap in general (Köhler et al., 2012; McKeown et al., 2011; Pignatta & Gehring, 2012), described conserved imprinted expression of three genes only (Pignatta and Gehring, 2012; Jiang and Köhler, 2012). This indicates that in monocots (rice, maize) and dicots (*Arabidopsis*), which diverged about 140-150 million years ago (Chaw et al., 2004), the selective advantage was strong enough to maintain imprinting over such long period of time. Therefore, studying genomic imprinting in plants offers the great opportunity to combine natural variation, evolution, and epigenetic gene regulation.

***De novo* expression or carry-over?**

A weak point of many recent studies, including our own (Chapter 2.3), that should be considered more in the future, is the analysis of steady-state RNA levels. Therefore, it cannot be inferred that the analyzed transcripts are *de novo* expressed in the fertilization products rather than deposited in the gametes. Certainly, absence of a transcript in the transcriptome of the gametes suggests that the gene is *de novo* expressed in embryo and/or endosperm. Yet, we lack the required depth in gamete transcriptome analysis for egg cell and sperm cells (Wuest et al., 2010; Borges et al., 2008), but not for central cell (Schmid et al., 2012). Likely, the restricted access to and the difficult isolation of (female) plant gametes is a major obstacle to in-depth analysis of the gametic transcriptome or even the epigenome. Furthermore, alternative methods to show *de novo* transcription like nascent RNA fluorescence *in-situ* hybridization (RNA-FISH) are very difficult to perform in plants and it was successfully

applied only once to study *de novo* expression of maternal *MEA* allele in the fertilized central cell (Vielle-Calzada et al., 1999).

Generally, technological advancement in plant research is necessary and will hopefully solve a few of the current unknown aspects in the field of plant developmental genetics, e.g. sufficiently deep transcriptome or even epigenome data of plant stem cells and plant gametes. First, cell-type specific analysis in plants is lagging behind but is essential to any aspect of molecular biology. Current approaches include fluorescence assisted cell sorting (FACS), which unfortunately does not seem to work for cell types that are too rare within a tissue (like the female gametes within the ovule) and, additionally, requires efficient production of protoplasts by digesting the plant cell wall. Further approaches include laser-assisted microdissection (Wuest et al., 2010, Schmid et al., 2012) or manual or semi-manual manipulation of a given tissue (Gehring et al., 2011; Raissig et al., in press; Waters et al., 2011; Xiang et al., 2011), both very time-consuming techniques and not appropriate for large-scale tissue isolation. A very elegant and promising technology to isolate specific nuclei is the INTACT system (Deal and Henikoff, 2011) that is based on affinity-labeled nuclei by expression of a transgenic, cell type-specific biotinylated nuclear envelope protein. Total nuclei are isolated from a heterogenous tissue and specific nuclei can then be purified by streptavidin-coated magnetic beads (Deal & Henikoff, 2011). Hopefully, techniques like the INTACT system will be further developed to eventually isolate whole cells, for example by expressing biotinylated cell-membrane proteins.

Evolution of genomic imprinting: Combinatorial evolutionary driving forces or parental conflict?

Many evolutionary hypotheses have been proposed to explain the evolution of imprinting. To formulate evolutionary concepts, one has to take all available empirical data into account and not just develop mathematical models for the sake of mathematical modeling (Moore and Mills, 2008). The intragenomic parental conflict over resource allocation provides a widely accepted and testable theory for the evolution of imprinting, suggesting that growth restricting factors should be under maternal control to antagonize paternally expressed growth promoting factors (Moore and Haig, 1991; Haig and Westoby, 1989).

In both placental mammals and flowering plants, embryo development mostly depends on maternal resources. The parental conflict hypothesis is supported by a variety of experimental evidence: First, the placenta and the endosperm are the primary tissues exhibiting genomic imprinting (Frost & Moore, 2010; Jullien & Berger, 2009; Raissig et al., 2011). Second, mutations in many imprinted genes in animals and plants display growth and developmental defects (Angiolini et al., 2006; Raissig et al., 2011; Reik et al., 2003; Tycko & Morison, 2002). Third, genomic imprinting seems to be absent from egg-laying mammals, the monotremes, suggesting the concurrent evolution of genomic imprinting and the "placental habit" (Suzuki et al., 2007; Edwards et al., 2008).

However, empirical data challenge the parental conflict theory: For example the imprinted gene *Meg1* in maize is maternally expressed, is required for the development of transfer cells between mother and offspring, and is involved in nutrient transfer (Costa et al., 2012). RNAi mutants of *Meg1* produce smaller seeds, a phenotype that is not predicted by the parental conflict theory (Costa et al., 2012). The authors propose that this observation rather fits the co-adaptation theory, which predicts maternal expression of traits important at the mother-offspring interface. Furthermore, 89 MEGs and 61 PEGs are described in mammals (www.mousebook.org), whereas many more MEG candidates than PEG candidates were identified in the recent studies in *Arabidopsis* (290 MEGs vs. 79 PEGs; Wolff et al., 2011; Hsieh et al., 2011; Gehring et al., 2011; McKeown et al., 2011) and rice (177 MEGs vs. 85 PEGs; Luo et al., 2011) but not in maize (122 MEGs vs. 157 PEGs; Zhang et al., 2011; Waters et al., 2011). The prevalence of MEGs compared to PEGs is even more dramatic considering the imprinted candidates in the

embryo of *Arabidopsis*, rice, and also maize (Hsieh et al., 2011; Gehring et al., 2011; Luo et al., 2011; Waters et al., 2011; Chapter 2.3, Table D1). In fact, the co-adaptation theory predicts many more MEGs than PEGs and could explain the higher occurrence of MEGs (Wolf & Hager, 2006). However, the plant embryo and endosperm are surrounded by the maternal seed coat, and even the endosperm is triploid, containing two maternal and one paternal genome. Therefore, the MEG bias described in the recent profiling studies might be an artifact and could be due to the special tissue composition of plant seeds, which can be of maternal origin, like the seed coat, or maternally dominated, like the endosperm, and are hard to separate. In conclusion, before the candidates are not confirmed by alternative methods or called more than once in different studies, this observation not necessarily contradicts the parental conflict theory.

However, the question remains how we can reconcile conflicting data to fit a single theory in the future. Einstein once said, "If the facts don't fit the theory change the facts" and, therefore, proposes to independently reassess ill-fitting facts, like the observed MEG bias in plants. In contrast, it could be considered that different driving forces are responsible for the evolution of genomic imprinting and that specific combinations caused genomic imprinting at one locus or in one tissue, whereas other combinations underlie the evolution of other imprinted loci or in other tissues. We believe that studying the function of recently identified imprinted genes in plants will aid to understand the evolution of imprinted traits. A difficult aspect is the dosage-sensitivity of imprinted genes, whereby full ablation of the gene function might not necessarily give the right answer but rather requires fine manipulation of gene dosage. Mutants in the recently identified MEGs and the PEG in the *Arabidopsis* embryo (Chapter 2.3, Raissig et al. in preparation) do not show an obvious fertility or embryo patterning phenotype. However, many of the embryonic MEGs are involved in metabolism, potentially fine-tuning embryonic growth and development. To overcome potential gene redundancy, which obviously would mask loss-of-function mutant, on could rather increase gene dose by overexpression constructs and assess for example seed or embryo growth rates and/or size. Taken together, functional analyses, which are easier and cheaper to perform in plants than in the mouse, where the creation of transgenic individuals is time-consuming and costly, will shed light on the function and evolution of genomic imprinting.

Genomic imprinting in plants – How and where?

The main research goals of this thesis were to elucidate the mechanisms that regulate genomic imprinting in *Arabidopsis* (how?), and to investigate whether imprinting in *Arabidopsis* is really restricted to the endosperm (where?). We could demonstrate that parent-of-origin monoallelic gene expression is indeed a feature of the plant embryo and, therefore, overcome the dogma that the endosperm is the sole tissue exhibiting imprinting in plants. In addition, we showed that genes could be imprinted in one tissue although they are not in others (lineage-specific imprinting), and that imprinting at some loci is dynamically regulated in the course of development. Unfortunately, we were less lucky in identifying *trans*-acting imprinting regulators acting at the *MEA*-ICR in an allele-specific manner. However, we were able to identify two maternal activator candidates that are required for seed development but could not map the mutations by SRM, likely due to the FUKUSHIMA (FUK) background.

In conclusion, future research accompanied by technical advancements in large-scale and specific tissue isolation, accompanied by ever dropping costs to use next-generation-sequencing platforms for ChIP-Seq and allele-specific RNA-Seq, will surely help to clarify the plethora of unanswered aspects of plant genomic imprinting including regulatory mechanisms, function and evolution, and lineage-specificity of imprinted gene expression.

References:

Angiolini, E., Fowden, A., Coan, P., Sandovici, I., Smith, P., Dean, W., Burton, G., Tycko, B., Reik, W., Sibley, C., and Constância, M. (2006). Regulation of placental efficiency for nutrient transport by imprinted genes. Placenta **27 Suppl A**: S98–102.

Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., Jiang, R., Muliyati, N.W., Zhang, X., Amer, M.A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., De Meaux, J., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature **465**: 627–631.

Austin, R.S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N.J., Bonetta, D., Zhang, J., Fung, P., Gong, Y., Wang, P.W., McCourt, P., and Guttman, D.S. (2011). Next-generation mapping of *Arabidopsis* genes. Plant J **67**: 715–725.

Autran, D., Baroux, C., Raissig, M.T., Lenormand, T., Wittig, M., Grob, S., Steimer, A., Barann, M., Klostermeier, U.C., Leblanc, O., Vielle-Calzada, J.-P., Rosenstiel, P., Grimanelli, D., and Grossniklaus, U. (2011). Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. Cell **145**: 707–719.

Barlow, D.P. (2011). Genomic imprinting: a Mammalian epigenetic discovery model. Ann Rev Genet **45**: 379–403.

Baroux, C., Autran, D., Gillmor, C.S., Grimanelli, D., and Grossniklaus, U. (2009). The maternal to zygotic transition in animals and plants. Cold Spring Harbor Symp Quant Biol **73**: 89-100.

Bell, A.C. and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. Nature **405**: 482–485.

Benoit, P., Papin, C., Kwak, J.E., Wickens, M., and Simonelig, M. (2008). PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in *Drosophila*. Development **135**: 1969–79.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J.A., and Becker, J.D. (2008). Comparative transcriptomics of *Arabidopsis* sperm cells. Plant Phys **148**: 1168–1181.

Chaw, S.-M., Chang, C.-C., Chen, H.-L., and Li, W.-H. (2004). Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. J Mol Evol **58**: 424–441.

Chen, Z.J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci **15**: 57–71.

Chotalia, M., Smallwood, S.A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., James, K., Dean, W., and Kelsey, G. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev **23**: 105–117.

Ciccone, D.N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., and Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature **461**: 415–418.

Costa, L.M., Yuan, J., Rouster, J., Paul, W., Dickinson, H., and Gutierrez-Marcos, J.F. (2012). Maternal control of nutrient allocation in plant seeds by genomic imprinting. Curr Biol **22**: 160–165.

Deal, R.B. and Henikoff, S. (2011). The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat Protoc **6**: 56–68.

Deveale, B., Van Der Kooy, D., and Babak, T. (2012). Critical Evaluation of Imprinted Gene Expression by RNA–Seq: A New Perspective. PLoS Genet **8**: e1002600.

Edwards, C.A., Mungall, A.J., Matthews, L., Ryder, E., Gray, D.J., Pask, A.J., Shaw, G., Graves, J.A.M., Rogers, J., consortium, S., Dunham, I., Renfree, M.B., and Ferguson-Smith, A.C. (2008). The evolution of the DLK1-DIO3 imprinted domain in mammals. Plos Biol **6**: e135.

Ferrón, S.R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., Morante-Redolat, J.M., Laborda, J., Guillemot, F., Bauer, S.R., Fariñas, I., and Ferguson-Smith, A.C. (2011). Postnatal loss of *Dlk1* imprinting in stem cells and niche astrocytes regulates neurogenesis. Nature **475**: 381–385.

Frost, J.M. and Moore, G.E. (2010). The importance of imprinting in the human placenta. PLoS Genet **6**: e1001015.

Galili, G., Kawata, E.E., Smith, L.D., and Larkins, B.A. (1988). Role of the 3'-poly(A) sequence in translational regulation of mRNAs in *Xenopus laevis* oocytes. J Biol Chem **263**: 5764–70.

Gazzani, S., Gendall, A.R., Lister, C., and Dean, C. (2003). Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. Plant Phys **132**: 1107–14.

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic Analysis of Parent-of-Origin Allelic Expression in *Arabidopsis thaliana* Seeds. PLoS ONE **6**: e23687.

Glass, J.L., Fazzari, M.J., Ferguson-Smith, A.C., and Greally, J.M. (2009). CG dinucleotide periodicities recognized by the Dnmt3a-Dnmt3L complex are distinctive at retroelements and imprinted domains. Mamm Genome **20**: 633–643.

Gutiérrez-Marcos, J.F., Costa, L.M., Dal Prà, M., Scholten, S., Kranz, E., Perez, P., and Dickinson, H.G. (2006). Epigenetic asymmetry of imprinted genes in plant gametes. Nat Genet **38**: 876–878.

Haig, D. and Westoby, M. (1989). Parent-Specific Gene Expression and the triploid Endosperm. American Naturalist **134**: 147–155.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. Nature **405**: 486–489.

Haun, W.J. and Springer, N.M. (2008). Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. Plant J **56**: 903–912.

He, G., Zhu, X., Elling, A.A., Chen, L., Wang, X., Guo, L., Liang, M., He, H., Zhang, H., Chen, F., Qi, Y., Chen, R., and Deng, X.-W. (2010). Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell **22**: 17–33.

Heo, J.B. and Sung, S. (2010). Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. Science **331**: 76-79.

Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., and Fischer, R.L. (2011). Inaugural Article: Regulation of imprinted gene expression in *Arabidopsis* endosperm. Proc Natl Acad Sci USA **108**: 1755–1762.

Hudson, Q.J., Kulinski, T.M., Huetter, S.P., and Barlow, D.P. (2010). Genomic imprinting mechanisms in embryonic and extraembryonic mouse tissues. Heredity **105**: 45–56.

Immink, R.G.H., Posé, D., Ferrario, S., Ott, F., Kaufmann, K., Valentim, F.L., De Folter, S., Van der Wal, F., Van Dijk, A.D.J., Schmid, M., and Angenent, G.C. (2012). Characterization of SOC1's central role in flowering by the identification of its upstream and downstream regulators. Plant Phys **160**: 433–449.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature **416**: 556–60.

Jahnke, S. and Scholten, S. (2009). Epigenetic resetting of a gene imprinted in plant embryos. Curr Biol **19**: 1677–1681.

Jiang, H. and Köhler, C. (2012). Evolution, function, and regulation of genomic imprinting in plant seed development. J Exp Bot **63**: 4713–4722.

Jullien, P.E. and Berger, F. (2009). Gamete-specific epigenetic mechanisms shape genomic imprinting. Curr Op Plant Biol **12**: 637–642.

Kermicle, J.L. (1970). Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. Genetics **66**: 69–85.

Kermicle, J.L. and Alleman, M. (1990). Gametic imprinting in maize in relation to the angiosperm life cycle. Dev Suppl **1990**: 9-14.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. Science **303**: 521–523.

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J., Koornneef, M., and Kakutani, T. (2007). Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. Plant J **49**: 38–45.

Köhler, C., Wolff, P., and Spillane, C. (2012). Epigenetic mechanisms underlying genomic imprinting in plants. Ann Rev Plant Biol **63**: 331–352.

Lasko, P. (2009). Translational control during early development. Prog Mol Biol Transl Sci **90**: 211–54.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. Dev Cell **15**: 547–557.

Lindner, H., Raissig, M., Sailer, C., Shimosato-Asano, H., Bruggmann, R., and Grossniklaus, U. (2012). SNP-Ratio Mapping (SRM): Identifying Lethal Alleles and Mutations in Complex Genetic Backgrounds by Next-Generation Sequencing. Genetics **191**: 1381–1386.

Luo, M., Taylor, J., Spriggs, A., Zhang, H., Wu, X., Russell, S., Singh, M., and Koltunow, A. (2011). A Genome-Wide Survey of Imprinted Genes in Rice Seeds Reveals Imprinting Primarily Occurs in the Endosperm. PLoS Genet **7**: e1002125.

Lyle, R., Watanabe, D., Te Vruchte, D., Lerchner, W., Smrzka, O.W., Wutz, A., Schageman, J., Hahner, L., Davies, C., and Barlow, D.P. (2000). The imprinted antisense RNA at the *Igf2r* locus overlaps but does not imprint *Mas1*. Nat Genet **25**: 19–21.

Mancini-Dinardo, D., Steele, S.J.S., Ingram, R.S., and Tilghman, S.M. (2003). A differentially methylated region within the gene *Kcnq1* functions as an imprinted promoter and silencer. Hum Mol Genet **12**: 283–294.

Martienssen, R.A. (2010). Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of *Arabidopsis*. New Phytol **186**: 46–53.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009). RNA-mediated chromatin-based silencing in plants. Curr Op Cell Biol **21**: 367–376.

Mayer, U., Büttner, G., and Jürgens, G. (1993). Apical-basal pattern formation in the *Arabidopsis* embryo: studies on the role of the *gnom* gene. Development **117**: 149–162.

McKeown, P.C., Laouielle-Duprat, S., Prins, P., Wolff, P., Schmid, M.W., Donoghue, M.T., Fort, A., Duszynska, D., Comte, A., Lao, N.T., Wennblom, T.J., Smant, G., Köhler, C., Grossniklaus, U., and Spillane, C. (2011). Identification of imprinted genes subject to parent-of-origin specific expression in *Arabidopsis thaliana* seeds. BMC Plant Biol **11**: 113.

Messing, J. and Grossniklaus, U. (1999). Genomic imprinting in plants. Results Probl Cell Differ **25**: 23–40.

Moore, T. and Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet **7**: 45–49.

Moore, T. and Mills, W. (2008). Evolutionary theories of imprinting--enough already! Adv Exp Med Biol **626**: 116–122.

Nodine, M.D. and Bartel, D.P. (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature **482**: 94–97.

Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N.A., Shah, C., Wall, J.D., Wang, J., Zhao, K., et al. (2005). The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol **3**: e196.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., Cheng, X., and Bestor, T.H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. Nature **448**: 714–717.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell **32**: 232–246.

Pignatta, D. and Gehring, M. (2012). Imprinting meets genomics: new insights and new challenges. Curr Opin Plant Biol **15**: 530-535.

Raissig, M., Gagliardini, V., Jaenisch, J., Grossniklaus, U., and Baroux, C. Efficient and rapid isolation of early-stage embryos from *Arabidopsis thaliana* seeds. J Vis Exp - *in press*

Raissig, M.T., Baroux, C., and Grossniklaus, U. (2011). Regulation and Flexibility of Genomic Imprinting during Seed Development. Plant Cell **23**: 16–26.

Redrup, L., Branco, M.R., Perdeaux, E.R., Krueger, C., Lewis, A., Santos, F., Nagano, T., Cobb, B.S., Fraser, P., and Reik, W. (2009). The long noncoding RNA *Kcnq1ot1* organises a lineage-specific nuclear domain for epigenetic gene silencing. Development **136**: 525–530.

Reik, W., Constância, M., Fowden, A., Anderson, N., Dean, W., Ferguson-Smith, A., Tycko, B., and Sibley, C. (2003). Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. J Physiol **547**: 35–44.

Riddle, N.C. and Richards, E.J. (2002). The control of natural variation in cytosine methylation in *Arabidopsis*. Genetics **162**: 355–63.

Rojo, E., Gillmor, C.S., Kovaleva, V., Somerville, C.R., and Raikhel, N. V (2001). *VACUOLELESS1* is an essential gene required for vacuole formation and morphogenesis in *Arabidopsis*. Dev Cell **1**: 303–310.

Santoro, F. and Barlow, D.P. (2011). Developmental control of imprinted expression by macro non-coding RNAs. Semin Cell Dev Biol **22**: 328–335.

Saze, H. and Kakutani, T. (2007). Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. EMBO J **26**: 3641–52.

Schmid, M.W., Schmidt, A., Klostermeier, U.C., Barann, M., Rosenstiel, P., and Grossniklaus, U. (2012). A powerful method for transcriptional profiling of specific cell types in eukaryotes: laser-assisted microdissection and RNA sequencing. PLoS ONE **7**: e29685.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jørgensen, J.E., Weigel, D., and Andersen, S.U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods **6**: 550–551.

Springer, N.M. and Stupar, R.M. (2007). Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. Plant Cell **19**: 2391–2402.

Suzuki, S., Ono, R., Narita, T., Pask, A.J., Shaw, G., Wang, C., Kohda, T., Alsop, A.E., Marshall Graves, J.A., Kohara, Y., Ishino, F., Renfree, M.B., and Kaneko-Ishino, T. (2007). Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. PLoS Genet **3**: e55.

Szabó, P., Tang, S.H., Rentsendorj, A., Pfeifer, G.P., and Mann, J.R. (2000). Maternal-specific footprints at putative CTCF sites in the *H19* imprinting control region give evidence for insulator function. Curr Biol **10**: 607–610.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., Van Lohuizen, M., Orkin, S.H., and Peters, A.H.F.M. (2008). *Polycomb* group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev Cell **15**: 668–679.

Tessadori, F., Van Zanten, M., Pavlova, P., Clifton, R., Pontvianne, F., Snoek, L.B., Millenaar, F.F., Schulkes, R.K., Van Driel, R., Voesenek, L.A.C.J., Spillane, C., Pikaard, C.S., Fransz, P., and Peeters, A.J.M. (2009). Phytochrome B and histone deacetylase 6 control light-induced chromatin compaction in *Arabidopsis thaliana*. PLoS Genet **5**: e1000638.

Tycko, B. and Morison, I.M. (2002). Physiological functions of imprinted genes. J Cell Physiol **192**: 245–258.

Vielle-Calzada, J.P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M.A., and Grossniklaus, U. (1999). Maintenance of genomic imprinting

at the *Arabidopsis MEDEA* locus requires zygotic *DDM1* activity. Genes Dev **13**: 2971–2982.

Villar, C.B.R. and Köhler, C. (2010). Plant chromatin immunoprecipitation. Methods Mol Biol **655**: 401–411.

Waters, A.J., Makarevitch, I., Eichten, S.R., Swanson-Wagner, R.A., Yeh, C.T., Xu, W., Schnable, P.S., Vaughn, M.W., Gehring, M., and Springer, N.M. (2011). Parent-of-Origin Effects on Gene Expression and DNA Methylation in the Maize Endosperm. Plant Cell **23**: 4221–4233.

Wolf, J.B. and Hager, R. (2006). A maternal-offspring coadaptation theory for the evolution of genomic imprinting. PLoS Biol **4**: e380.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C. (2011). High-Resolution Analysis of Parent-of-Origin Allelic Expression in the *Arabidopsis* Endosperm. PLoS Genet **7**: e1002126.

Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenführer, J., Von Mering, C., and Grossniklaus, U. (2010). *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. Curr Biol **20**: 506–512.

Wutz, A., Smrzka, O.W., Schweifer, N., Schellander, K., Wagner, E.F., and Barlow, D.P. (1997). Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. Nature **389**: 745–749.

Wöhrmann, H.J., Gagliardini, V., Raissig, M.T., Wehrle, W., Arand, J., Schmidt, A., Tierling, S., Page, D.R., Schöb, H., Walter, J., and Grossniklaus, U. (2012). Identification of a DNA methylation-independent imprinting control region at the *Arabidopsis MEDEA* locus. Genes Dev **26**: 1837–1850.

Xiang, D., Venglat, P., Tibiche, C., Yang, H., Risseeuw, E., Cao, Y., Babic, V., Cloutier, M., Keller, W., Wang, E., Selvaraj, G., and Datla, R. (2011). Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*. Plant Phys **156**: 346–356.

Zhang, X. and Borevitz, J.O. (2009). Global analysis of allele-specific expression in *Arabidopsis thaliana*. Genetics **182**: 943–954.

Zhong, X., Hale, C.J., Law, J.A., Johnson, L.M., Feng, S., Tu, A., and Jacobsen, S.E. (2012). DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. Nat Struct Mol Biol **19**: 870–875.

Zuryn, S., Le Gras, S., Jamet, K., and Jarriault, S. (2010). A strategy for direct mapping and identification of mutations by whole-genome sequencing. Genetics **186**: 427–430.

APPENDIX

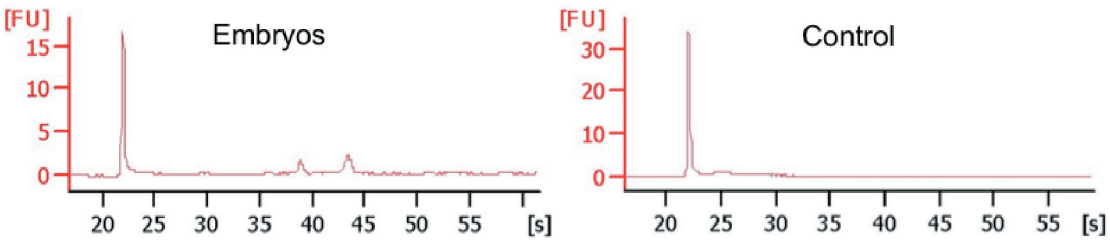
Supplemental Data.

Supplemental Figure 1. Embryo Isolation and RNA Extraction.

In recent years the expression of *MEA* in the embryo has become a matter of some debate, possibly because current models do not explain activation of *MEA* in the embryo and because certain *pMEA:GUS* reporter lines have been used as central cell-specfic markers (Gross-Hardt et al., 2007). It is important to recognize, however, that these *promoter:GUS* reporter lines do not necessarily accurately reflect endogenous expression patterns, particularly if some *cis*-regulatory regions are missing or position effects influence reporter gene expression. To substantiate previous reports from several laboratories, we performed GUS staining on isolated embryos and RT-PCR on RNA from isolated embryos, results of which are shown in Figure 1 of the main text. The *pMEA:GUS* reporter used for embryo staining covers the full *MEA* promoter and 1kb of the coding region fused to *GUS* (Baroux et al., 2006). This data confirms embryonic expression of *MEA*, as previously reported by a number of different research groups using RT-PCR on isolated embryos (Kinoshita et al., 1999; Gehring et al. 2006), reporter gene analysis (Luo et al., 2000; Baroux et al., 2006; Wang et al., 2006), and *in situ* hybridization (Vielle-Calzada et al., 1999; Spillane et al., 2007).

Here, we provide supplemental data for the results presented in Figure 1B, showing images of isolated embryos used for RNA extraction (Supplemental Figure 1A), and PicoChip bioanalyzer RNA profiles (Supplemental Figure 1B). For RT-PCR, a washing step was used to remove debris and potentially contaminating endosperm nuclei. Supplemental Figure 1B shows that no RNA was detected in samples with embryos removed (Control) compared to embryonic RNA detected with isolated embryos present (Embryos).

Supplemental Figure 1A. Isolated embryos representative of those used for RNA extraction.



Supplemental Figure 1B. PicoChip bioanalyzer RNA profiles from isolated embryo samples.

Supplemental References

Baroux, C., Gagliardini, V., Page, D.R., and Grossniklaus, U. (2006). Dynamic regulatory interactions of *Polycomb* group genes: *MEDEA* autoregulation is required for imprinted gene expression in *Arabidopsis*. Genes Dev **20:** 1081-1086.

Gehring, M., Huh, J.H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes *MEDEA Polycomb* gene self-imprinting by allele-specific demethylation. Cell **124:** 495-506.

Gross-Hardt, R., Kägi, C., Baumann, N., Moore, J.M., Baskar, R., Gagliano, W.B., Jürgens, G., Grossniklaus, U. (2007). *LACHESIS* restricts gametic cell fate in the female gametophyte of *Arabidopsis*. PLoS Biol **5**: e47.

Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Imprinting of the *MEDEA Polycomb* gene in the *Arabidopsis* endosperm. Plant Cell **11:** 1945-1952.

Luo, M., Bilodeau, P., Dennis, E.S., Peacock, W.J., and Chaudhury, A. (2000). Expression and parent-of-origin effects for *FIS2*, *MEA*, and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. Proc Natl Acad Sci USA **97:** 10637-10642.

Spillane, C., Schmid, K.J., Laoueillé-Duprat, S., Pien, S., Escobar-Restrepo, J.-M., Baroux, C., Gagliardini, V., Page, D.R., Wolfe, K.H., and Grossniklaus, U. (2007). Positive darwinian selection at the imprinted *MEDEA* locus in plants. Nature **448:** 349-352.

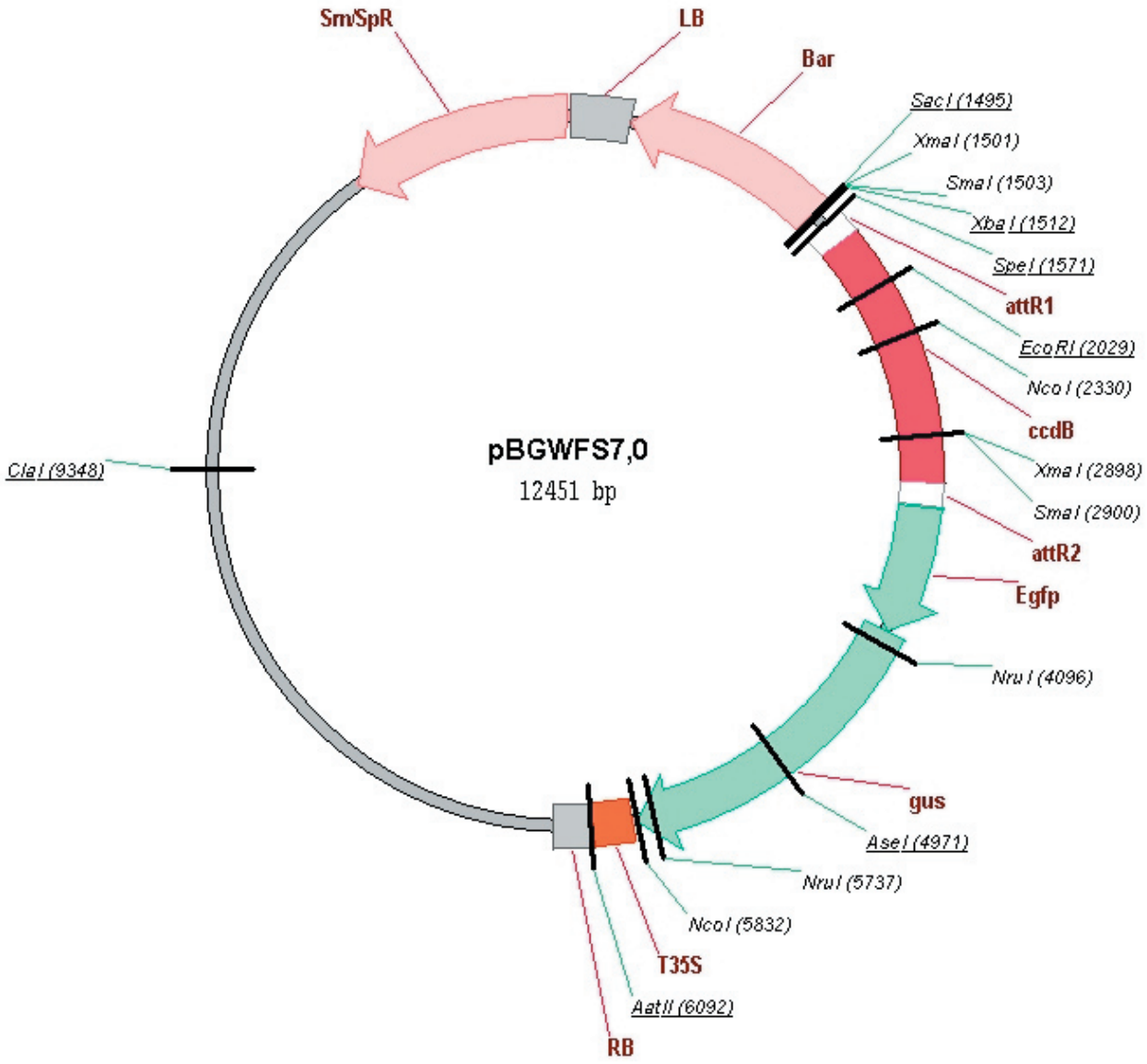
Vielle-Calzada, J.P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M.A., and Grossniklaus, U. (1999). Maintenance of genomic imprinting at the *Arabidopsis MEDEA* locus requires zygotic *DDM1* activity. Genes Dev **13:** 2971-2982.

Wang, D., Tyson, M.D., Jackson, S.S., and Yadegari, R. (2006). Partially redundant functions of two SET-domain *Polycomb* group proteins in controlling initiation of seed development in *Arabidopsis*. Proc Natl Acad Sci U S A. **103**: 13244-13249.

PRIMER LIST

MRprimer 1	GCTGTCAAACGTCAAGCAAA	51°	Amplifying 250bp minimal MEA promoter
MRprimer 2	AACCACTCGCCTCTTCTTTTT	51°	Amplifying 250bp minimal MEA promoter
MRprimer 3	CACCAAGCTCGCACTTCTT	?	Genotpying hac10-2 forward
MRprimer 4	CGAATTAGATCAACTGCTGAGG	?	Genotpying hac10-2 reverse
priMR 5	acgactcgtccgtcctgtag	?	amplifying GUS probe for Southern blots
priMR 6	tgatgataatcggctgatgc	?	amplifying GUS probe for Southern blots
priMR 7	CAGTGGTCTCCTTGATGATGC	?	fas2-4_f
priMR 8	GAGAGGCTACATACTTTGCTAA	?	fas2-4_r
priMR 9	GAGATAATGACGACGAAGATGGTGC	60°C, 20sec	250pMEA genotyping: wt & T-DNA (SGP30)
priMR 10	ATCTCCATCTCTTCCAGTAGCAGCC	60°C, 20sec	250pMEA genotyping: wt (SGP31)
priMR 11	TAGAGTCCCGCAATTATACATTTAATACGC	58°C, 20sec	250pMEA genotyping: T-DNA (RB2 von Stefan Grob, RB primer of pCAMBIA1381Z)
priMR 12	CGCTACCTACTGCTACTTTCAAGA	58°C, 20sec	4.8kbMEA genotyping: wt (SGP32)
priMR 13	ATGCCTAACCTGAAACTCAAACCAG	58°C, 20sec	4.8kbMEA genotyping: wt (SGP33)
priMR 14	CTATAGGGTTTCGCTCATGTGTTGAG	59°C, 20sec	4.8kbMEA genotyping: T-DNA (SGP23, LB primer of pCAMBIA 1381Z)
priMR 15	GTTAAACATGTTGGAATACCATTGAATAG	59°C, 20sec	4.8kbMEA genotyping: T-DNA (SGP63)
priMR 16	TGGCTTAAGTGTGAGACCTCC	59°(?), 20sec	FLAG_062A06 genotyping (LP), not really unique if blasted.. Tja...
priMR 17	TTTTCCCCAAAAATGATGTTG	59°(?), 20sec	FLAG_062A06 genotyping (RP), this one to be used with LB_FLAG-Primer for mutant band!!!
priMR 18	AACATCAATGGCGACTCCAG		dCAPS, chr.I, 25Mb, At1G66220, cuts wt forward with PvuII
priMR 19	GATTGCAGCTGGAGACCAGT		dCAPS, chr.I, 25Mb, At1G66220, cuts wt forward with PvuII
priMR 20	ATTTTGCCGATTTCGGAAC		LBb1.3, left border primer for genotyping SALK lines (gives a 200bp smaller mutant product, compared to LB1a!!!)
priMR 21	AAAATTTCTTCTTTCCAAAGAACG	59°	LP for genotyping SALK_070336
priMR 22	GTTTCTCGTCCCCACTTCTTC	60°	RP for genotyping SALK_070336 (wt-band = 1102, m-band = 498-798)
priMR 23	TGGTTGGATACGAACATTTCG	60°	LP for genotyping SALK_050279
priMR 24	TCACAGTCGTGACTCGTGAAG	60°	RP for genotyping SALK_050279 (wt-band = 1179, m-band = 500-880)
priMR 25	ACCAGTTTACAAACCCGAACC	60°	LP for genotyping SALK_065223.53.60
priMR 26	GGTGGAAGTGGAGTGTCAAAG	59°	RP for genotyping SALK_065223.53.60 (wt = 1048, m = 525-825)
priMR 27	CTTAGGGGTTTCCAATTTTGG	59°	LP for genotyping SALK_065223.40.10
priMR 28	AGCTTGGGAGTGTTGTTGTTG	60°	RP for genotyping SALK_065223.40.10 (wt = 1123, m = 549-849)
priMR 29	CTTAGGGGTTTCCAATTTTGG	59°	LP for genotyping SALK_065340.40.10
priMR 30	AGCTTGGGAGTGTTGTTGTTG	60°	RP for genotyping SALK_065340.40.10 (wt = 1123, m = 549-849)
priMR 31	CGATTTTCTTCAAGAGTTGCG	60°	LP for genotyping SALK_002322.44.30.x
priMR 32	CAAACCCATGAATTGGTTACAC	59°	RP for genotyping SALK_002322.44.30.x (wt = 1092, m = 437-737)
priMR 33	TGTTTGTGTGACGAGAAGACG	60°	LP for genotyping SALK_098050.48.85.x
priMR 34	CGTTCAGACCAGATTCTCTCG	60°	RP for genotyping SALK_098050.48.85.x (wt = 1028, m = 473-773)
priMR 35	TCCTTTCCAAGTCCCAAATTC	60°	LP for genotyping SALK_017591.30.60.x
priMR 36	ACTCCCACCACAACAAATACG	59°	RP for genotyping SALK_017591.30.60.x (wt = 1126, m = 513 - 813)
priMR 37	TTACTAAGACAAACACACCC		different reverse primer, replacing priMR 19 (shorter)
priMR 38	CCGCTCTTGTCACAACAGGT		different reverse primer, replacing priMR 19 (longer)
priMR 39	GGCTGACTGTGACAAAGAGGA		LP for SURVEYOR assay, At3G29638
priMR 40	AAAAAGTCcttccccgatct		RP for SURVEYOR assay, At3G29638
priMR 41	CTTCCTCGCTTAGCATCACAC	60°	LP for genotyping SALK_039610, At3G29638
priMR 42	TTGATTCACTTACGCAGGAGC	60°	RP for genotyping SALK_039610, At3G29638

VECTOR MAP - pBGWFS7



Supplemental Information

SUPPLEMENTAL MATERIALS AND METHODS

Plant Material

Wild-type plants were *Arabidopsis thaliana* accession Landsberg (with the *erecta* mutation, L*er*). Plants were cultured in a growth chamber at 70% humidity and daily cycles of 16h light at 21ºC and 8h darkness at 18ºC. For crosses, flowers were emasculated 2 days before pollination. Siliques were harvested at various times after pollination, as indicted in figure legends and in the text.

The mutant alleles used were *mea-1*, *mea-2* (L*er*) (Grossniklaus et al. 1998), *dme-4* (C24) (Guitton et al. 2004), and *met1-3* (Col) (Saze et al. 2003), and the *4.8pMEA::GUS* transgenic line (Spillane et al. 2004). The *mea-1* and *mea-2* alleles are identical with respect to genetic behavior, phenotype, and seed abortion frequency (Grossniklaus et al. 1998). The *mea-1* allele carries a *Ds* element in the coding sequence, *mea-2* is the result of a remobilization of this *Ds* leaving a 7 bp footprint that disrupts the gene by introducing 2 stop codons (Grossniklaus et al. 1998). We used the *mea2* allele for qPCR experiments because it is expressed at the same level as the wild-type *MEA* allele but the footprint facilitated the design of allele-specific primers for qPCR. Homozygous *mea-1* and *mea-2* mutant plants were obtained by embryo rescue (described in (Vielle-Calzada et al. 1999)) and exclusively used in the F1 generation to avoid ectopic epigenetic mis-regulation of direct and indirect target genes (U. Grossniklaus and C. Spillane, unpublished). The *dme-4* (C24) and the *met1-3* (Col-0) mutants were introgressed into the L*er* background by crossing them at least 5 times as pollen parents to L*er* wild-type plants (≥97% L*er* background). The *met1-3* mutant plants were propagated as heterozygotes and only used for experiments when showing full methylation at the 180 bp CEN-repeat (Martinez-Zapater et al. 1986) in Southern blots, indicating an unaltered epigenome.

The *mea-1* and *mea-2* mutant plants were selected on growth medium with 50 µg/ml kanamycin; *met1-3* mutants were selected on growth medium with 10 µg/ml phosphinothricin. In parallel, mutants were genotyped by PCR with the following primers: For *mea-1*, a wild-type (268 bp) and a mutant PCR fragment (350 bp) were amplified using the primers MEAS11, MEAretirev2, and Ds3.1. For *mea-2,* a wild-type PCR fragment (236 bp) was amplified using MEA2retiS1 and MEA2retiAS1, and a mutant PCR fragment (180 bp) was amplified using MEA2retiS2 and Ds5.1. For *dme-4* genotyping a wild-type PCR fragment (190 bp) and a mutant PCR fragment (160 bp) were amplified using HW25 and HW26. For *met1-3* genotyping a wild-type PCR fragment (400 bp) was amplified using MEF2 and MER2, and a

mutant PCR fragment (600 bp) was amplified using MEF1 and TL2. Primer sequences are given in Supplemental Table S6.

Generation of *pMEA::MEA* and *pMEA::GUS* constructs

pMEA::MEA *constructs*: All *pMEA::MEA* constructs were cloned into pCAMBIA3300 containing the corresponding *MEA* promoter sequence and the entire *MEA* open reading frame (ORF) amplified from genomic L*er* DNA. The *MEA* ORF was amplified as two PCR products, with primers DP248 and DP71 (+1 to +960), and with UA171 and UA8 (+878 to +4199). These PCR products were subcloned into pDRIVE (QIAGEN) and into pBluescriptSK (Stratagene), respectively, and liberated as *Nco*I-*Pst*I fragment (+1 to +920) and *Pst*I-*Sac*I fragment (+921 to +4199), respectively. The *MEA* promoter sequences were amplified with primers DP200 and CK176 (-1328 to +1), MEA5'8 and CK176 (-1104 to +1), MEA5'9 and CK176 (-889 to +1), MEA5'10 and CK176 (-667 to +1), and MEA5'11 and CK176 (-444 to +1). These PCR products were subcloned as *Eco*RI-*Nco*I fragment (-1328 to +1) or *Bam*HI-*Nco*I fragments (all other PCR products) into pDRIVE and liberated as *Eco*RI-*Nco*I fragments. The *Eco*RI-*Sac*I opened pCAMBIA3300 served as vector for the corresponding *MEA* promoter fragments (*Eco*RI-*Nco*I) and the two *MEA* ORF fragments (*Nco*I-*Pst*I and *Pst*I-*Sac*I) in the subsequent ligation. The *1330pMEA::MEA* construct served as template for the shorter constructs. *MEA* promoter fragments were amplified from *1330pMEA::MEA* with primers MEA5'14 and DP71 (-369 to +960), DP202 and DP71 (-254 to +960), and MEA5'15 and DP71 (-151 to +960). These PCR products were subcloned into pDRIVE, liberated as *Eco*RI-*Pst*I fragments and subsequently ligated with the *Pst*I-*Sac*I fragment (+921 to +4199) using the *Eco*RI-*Sac*I opened *1330pMEA::MEA* construct as vector. For the *200pMEA::MEA* construct, a *Sca*I-*Pst*I fragment (-209 to +921) from the *1330pMEA::MEA* construct was subcloned into pDRIVE, liberated as *Eco*RI-*Pst*I fragment, and subsequently ligated with the *Pst*I-*Sac*I fragment (+921 to +4199) using the *Eco*RI-*Pst*I opened *1330pMEA::MEA* construct as vector. Primer sequences are given in Supplemental Table S6.

pMEA::GUS *constructs*: All *pMEA::GUS* constructs contain the corresponding *MEA* promoter sequence amplified from genomic L*er* DNA and were cloned in frame to the *GUS* reporter gene in pCAMBIA 1381Z. The *MEA* promoter sequences were amplified with primers DP200 and CK176 (-1328 to +1), MEA5'8 and CK176 (-1104 to +1), MEA5'9 and CK176 (-889 to +1), MEA5'10 and CK176 (-667 to +1), and MEA5'11 and CK176 (-444 to +1). These PCR products were subcloned as *Eco*RI-*Nco*I fragment (-1328 to +1) or *Bam*HI-

*Nco*I fragments (all other PCR products) into pDRIVE and liberated as *Eco*RI-*Nco*I fragments. The *Eco*RI-*Nco*I opened pCAMBIA1381Z served as vector for the corresponding *MEA* promoter fragments (*Eco*RI-*Nco*I) in the subsequent ligation. The *1330pMEA::GUS* construct served as template for the shorter constructs. *MEA* promoter fragments were amplified from *1330pMEA::GUS* with primers MEA5'14 and CK176 (-369 to +1), DP202 and CK176 (-254 to +1), and MEA5'15 and CK176 (-151 to +1). These PCR products were subcloned into pDRIVE, liberated as *Eco*RI-*Nco*I fragments and subsequently ligated into the *Eco*RI-*Nco*I opened *1330pMEA::GUS* construct as vector. For the *200pMEA::GUS* construct, a *Sca*I-*Nco*I fragment (-209 to +1) from the *1330pMEA::GUS* construct was subcloned into pDRIVE, liberated and subsequently ligated into the *Sma*I-*Nco*I opened *1330pMEA::GUS* construct as vector. For the *D1330pMEA::GUS* construct (lacking the *MEA* promoter region from -200 to -150), a first *MEA* promoter fragment was liberated as *Eco*RI-*Sca*I fragment (-1328 to -209) from the *1330pMEA::GUS* construct. A second *MEA* promoter fragment was amplified with primers HW35 and CK176 (-151 to +1), subcloned into pDRIVE, and liberated as *Sca*I-*Nco*I fragment. The *Eco*RI-*Nco*I opened *1330pMEA::GUS* construct served as vector for the subsequent ligation of the *Eco*RI-*Sca*I fragment and *Sca*I-*Nco*I fragment, creating a 50bp deletion between -200 and -150 of the *MEA* promoter sequence. Primer sequences are given in Supplemental Table S6.

Microscopy and GUS staining

For histochemical analysis of GUS reporter gene expression, flowers were emasculated and collected after 48h (BF), or pollinated and collected 1-4 days after pollination (1-4 DAP). Unfertilized gynoecia and siliques were opened and directly transferred to the GUS reaction buffer for 48h-72h at 37ºC (2 mM 5-Bromo-4-chloro-3-indoxyl-beta-D-glucuronic acid, cyclohexylammonium salt (Biosynth AG, Staad, Switzerland), 10 mM EDTA, 0.1% Triton X-100, 2 mM potassium ferrocyanide, 2 mM potassium ferricyanide, 100 mM phosphate buffer, pH = 7.2). After incubation, unfertilized gynoecia and siliques 1-2 DAP were briefly rinsed in 100 mM phosphate buffer (pH = 7.2), cleared in 70% ethanol o/n, dissected and mounted in 80% glycerol; siliques 3-4 DAP were briefly rinsed in 100 mM phosphate buffer (pH = 7.2), cleared in 8:2:1 (w/w/w) chloral hydrate:water:glycerol o/n, dissected and mounted in the same solution.

Microscopic inspection of ovules and seeds was carried out under differential contrast (DIC) optics using a Leica DMR microscope (Leica Microsystems, Wetzlar, Germany).

Pictures were recorded using a Magnafire CCD camera (Optronics, Goleta, USA) and processed with Adobe Photoshop CS3 (Adobe Systems Inc., San Jose, USA).

RT-PCR Analyses

For RT-PCR analysis of *At1g02570* (Fig. S1), flowers before fertilization (BF) and approximately 1-3 DAP were collected. For *MEA* allele-specific RT-qPCR analysis in reciprocal crosses of *MEA/MEA* and *mea-2/mea-2* plants and in reciprocal crosses of *MEA/MEA; met1-3/MET* and *mea-2/mea-2* plants (Fig. 4), flowers were emasculated and the gynoecia collected 48h later (BF, before fertilization), or flowers were hand-pollinated 1 day after emasculation and harvested 1, 2, 3 or 4 DAP. Typically, 20 gynoecia or 10-15 siliques at 1-4 DAP were collected for RNA extraction. Tissues were immediately frozen in liquid nitrogen, ground, and total RNA was extracted using the NucleoSpin RNA Plant Kit (Machery-Nagel, Düren, Germany). Reverse transcription was performed as previously published (Baroux et al. 2006).

RT-PCRs for *MEA* and *At1g02570* were performed as follows: 4 min at 94ºC, 30 cycles (94ºC for 15 sec, 58ºC for 20 sec, and 72ºC for 30 sec) followed by 72ºC for 5 min. The RT-PCR for *Actin11* was performed with only 25 cycles and an annealing temperature of 60ºC. Primers used to amplify *MEA* were MEAS11 (F-primer) and MEAretirev2 (R-primer) spanning the 14th intron of *MEA*. Primers used to amplify *At1g02570* were HW45 (F-primer) and HW47 (R-primer) spanning the 1st intron of the gene. Primers used to amplify the control gene *ACTIN11* were Act11F (F-primer) and Act11R (R-primer). All primers are spanning introns, such that genomic DNA (gDNA)-specific and cDNA-specific RT-PCR products could easily be distinguished by size differences. Primer sequences are given in Supplemental Table S6.

RT-qPCR of *MEA* was performed as previously described (Baroux et al. 2006). Briefly, we used *MEA/MEA* and *mea-2/mea-2* homozygous plants to distinguish maternal and paternal transcripts; the *mea-2* mutant transcript contains a transposon sequence allowing specific detection and discrimination from the *MEA* wild-type transcripts (Grossniklaus et al. 1998). Using a Taqman RT-qPCR assay, we performed three PCR replicates for each sample. In all experiments, transcript levels were normalized to the level of *ACTIN11*.

Bisulfite DNA Sequencing of isolated reproductive cells

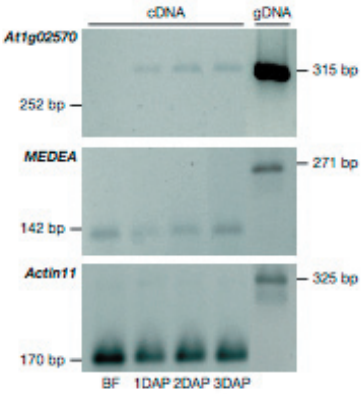
Central cells were isolated using LCM. The collected cells were digested using 1 volume lysis buffer (10 mM Tris-HCl, pH = 8, 10mM EDTA, 1 % SDS), 1 volume proteinase K (Roche,

Basel, Switzerland), and 100 ng salmon sperm DNA at 55 °C over night. Digested cells were bisulfite treated for 4 h at 50 °C including a first denaturing step for 15 min at 99 °C and two 5 min denaturing steps at 99 °C after 1 h and 2.5 h in 1.7 M sodium disulfite, 0.25 M NaOH both solved in H2O and 0.4 M 6- Hydroxy-2,5,7,8-tetramehtylchroman-2-carbonic-acid (Sigma-Aldrich, St. Luis, USA), solved in 1,4 Dioxan. Washing and desulfonation (10 min incubation in 0.3 M NaOH) steps followed on a microcon YM-30 column (Millipore, Billerica, USA). Bisulfite DNA was eluted in 50 µl TE. 5 µl were used for a 30 µl PCR reaction to amplify the MEA and FWA 5′ region (details see below) with HotStart-IT Taq DNA polymerase (USB, Cleveland Ohio, USA) and 454FusionPrimer P17 and P18 (Table S2). Purified PCR products were sequenced with the 454 sequencer according to the standard protocol.

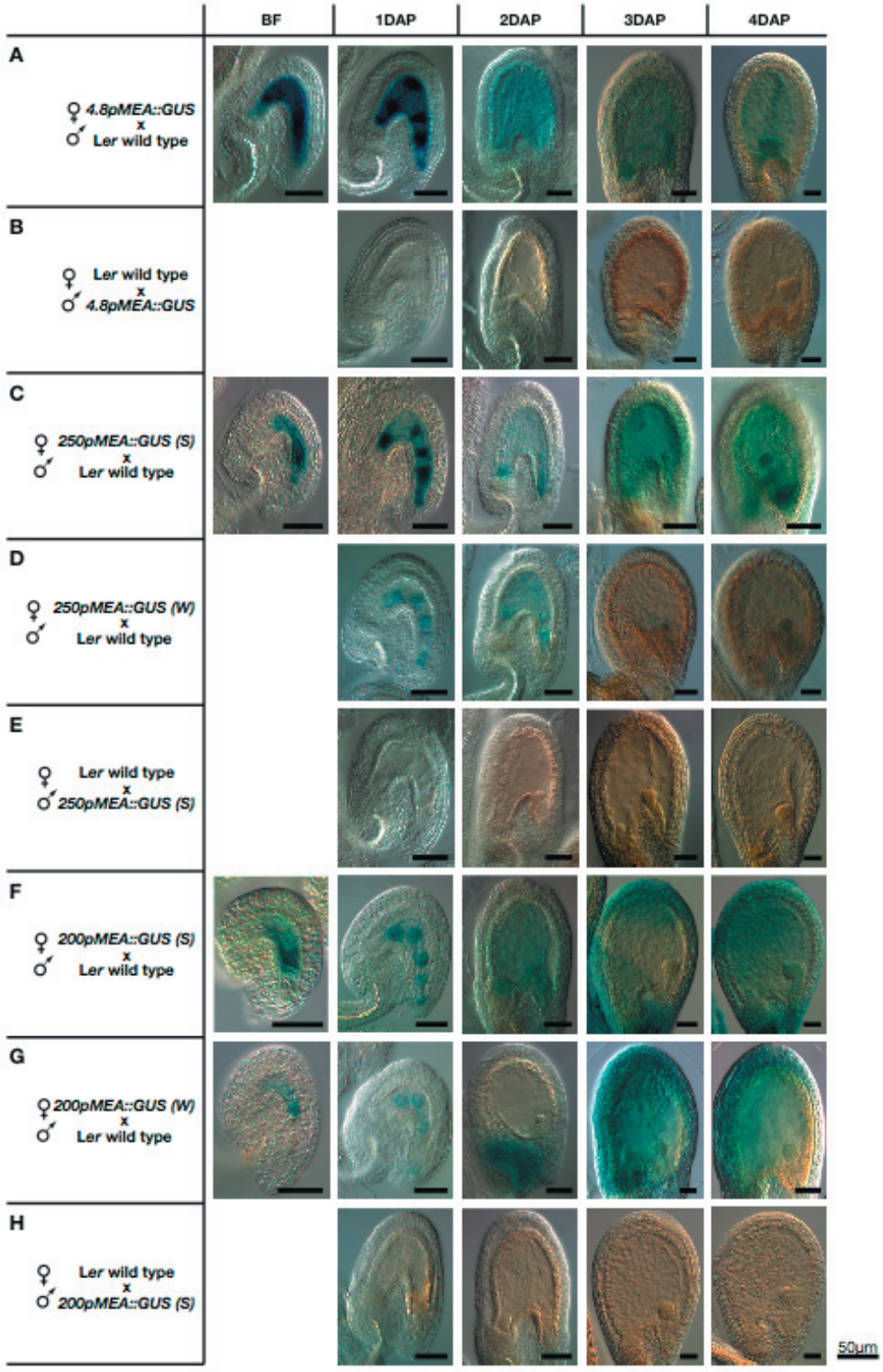
Sperm cells were isolated from ~20 mg L*er* wild-type pollen using a Percoll density gradient (M. Schauer and U. Grossniklaus, unpublished) and used for genomic DNA (gDNA) extraction with the QIAamp DNA Micro Kit (QIAGEN, Hilden Germany). Bisulfite conversion of ~50 µg sperm cell gDNA was performed with the EZ DNA Methylation–Direct Kit (Zymo Research, USA). Embryos (2-cell-stage, consisting of 2 embryo proper and 2 suspensor cells) were manually dissected from L*er* wild-type seeds as described in (Autran et al. 2011), released in standard TE-buffer, embedded in agarose, and subsequently treated with bisulfite (~100 embryos per sample). After bisulfite conversion the region covering the 250 bp *MEA* promoter (-342 to +115 relative to ATG) was amplified with primers MEDEA-A5-F and MEDEA-A5-R, and the region covering partially the SINE-related tandem repeat in the *FWA* promoter (-906 to -619 relative to ATG) was amplified with primers FWA-Bregion-F and FWA-Bregion-R under standard PCR conditions with the HotStarTaq DNA Polymerase (QIAGEN, Hilden, Germany). Purified bisulfite PCR products were cloned into the pGEM-T vector (Promega) and several independent clones were sequenced.

All sequences were analyzed with the BiQ analyzer software (Bock et al. 2005) for quality control and removal of identical clones in a standardized manner. Primer sequences are given in Supplemental Table S6. For a detailed description see the protocol “Bisulfite sequencing of small DNA/cell samples” (PROT35) at the Epigenetics Protocols Database (http://www.epigenome-noe.net/research tools/protocols.php).

Supplemental Figure S1



Supplemental Figure S2



Supplemental Figure Legends

Figure Legend S1. Expression analysis of *At1g02570*.
The RT-PCR indicates expression of *MEA*, but no expression of the 5'flanking gene *At1g02570*. RNA was isolated from unfertilized L*er*-0 wild-type siliques before fertilization (BF) and L*er*-0 wild-type siliques pollinated with *mea-2/mea-2* mutant pollen 1, 2 and 3 days after pollination (1-3DAP). *Actin11* expression was detected throughout early seed development and served as a control. Genomic DNA (gDNA) and cDNA amplification products are shown. DNA product sizes (in bp) are indicated along the sides of each panel.

Figure Legend S2. Expression analysis of *pMEA::GUS* transgenes during early seed development.
All *pMEA::GUS* transgenes were reciprocally crossed to L*er*-0 wild-type plants. GUS staining was monitored from before fertilization (BF) until 4 days after pollination (4 DAP).

Supplemental Tables

Table S1. Expression analysis of *pMEA::GUS* transgenes in independent T1 lines

pMEA::GUS	N staining	N total	N staining/ N total
Δ1330pMEA::GUS	0	142	0.00
1330pMEA::GUS	53	138	0.38
1100pMEA::GUS	11	29	0.38
890pMEA::GUS	10	24	0.42
670pMEA::GUS	11	27	0.41
450pMEA::GUS	8	25	0.32
370pMEA::GUS	18	55	0.33
250pMEA::GUS	19	116	0.16
200pMEA::GUS	15	104	0.14
150pMEA::GUS	0	105	0.00

The table shows the summary of the *pMEA::GUS* expression analysis in several independent primary transformants (T1). All T1 lines except for the ones with *D1330pMEA::GUS* and *150pMEA::GUS* show *MEA*-like GUS staining (compare **Fig.S2**). N staining designates the number of independent T1 lines showing *MEA*-like GUS staining. N total is the number of independent T1 lines analyzed for each *pMEA::GUS* transgene.

Table S2. Complementation analysis of *pMEA::MEA* transgenes in independent *mea/MEA* T1 lines

Rescue with *pMEA::MEA* aborted seeds	N full ≤25%	N partial 25% < N < 50%	N none 50%	N none >50%	N total	N rescue/ N total
1330pMEA::MEA	10	4	7	8	29	0.48
1100pMEA::MEA	30	5	9	2	46	0.76
890pMEA::MEA	10	1	9	0	20	0.55
670pMEA::MEA	29	5	13	2	49	0.69
450pMEA::MEA	28	8	9	3	48	0.75
370pMEA::MEA	35	3	9	1	48	0.79
250pMEA::MEA	10	4	7	2	23	0.61
200pMEA::MEA	20	0	14	1	35	0.57
150pMEA::MEA	0	0	30	2	32	0.00

The table shows the summary of the *pMEA::MEA* complementation analysis in several independent *mea-1/MEA* primary transformants (T1), which show 50% aborted seeds, unless the *pMEA::MEA* transgene is able to rescue the *mea*-induced seed abortion phenotype. We distinguished between full rescue with ≤ 25% aborted seeds, partial rescue with seed abortion between 25% and 50%, no rescue with 50% aborted seeds and no rescue with > 50% aborted seeds. N total is the number of independent T1 lines analyzed for each *pMEA::MEA* transgene. Assignment of the observed seed abortions in T1 lines to the four different seed abortion classes was based on a Chi-square best fit-scenario with $p < 0.05$.

Table S3. Imprinted genes with *MEA*-ICR-like sequences 3kb up- or downstream of the ORF

Potentially imprinted locus	Location of conserved regulatory sequence	p-value	Identity	GC-content of shared region	Reference
AT3G19160	**3' (+596bp until +755bp)**	**0.054**	**97/159 (61%)**	**0.21**	**Wolff et al. 2011**
AT2G19480	**5' (-815bp until -590bp)**	**0.23**	**124/225 (55%)**	**0.16**	**Gehring et al. 2011**
AT2G18880	**3' (+2514bp until +2792bp)**	**0.36**	**123/215 (57%)**	**0.18**	**Wolff et al. 2011**
AT1G62500	**3' (+1837bp until +1972bp)**	**0.49**	**83/135 (61%)**	**0.16**	**Gehring et al. 2011**
AT3G25250	**3' (+2003bp until +2240bp)**	**0.61**	**134/237 (56%)**	**0.15**	**Gehring et al. 2011**
AT3G43270	**5' (-540bp until -395bp)**	**0.76**	**86/145 (59%)**	**0.22**	**Gehring et al. 2011**
AT5G17320	3' (+902bp until +983bp)	0.998	56/81 (69%)	0.09	Gehring et al. 2009; 2011
AT2G05350	3' (+1593bp until +1669bp)	0.999	51/76 (67%)	0.06	Gehring et al. 2011
AT2G06050	5' (-1693bp until -1590bp)	0.999	65/103 (63%)	0.11	Gehring et al. 2011
AT3G01640	5' (-361bp until -270bp)	0.999	59/91 (64%)	0.22	Gehring et al. 2011
AT2G32750	5' (-391bp until -287bp)	0.9995	66/104 (63%)	0.15	Gehring et al. 2011
AT4G21430	3' (+2470bp until +2555bp)	0.9997	54/85 (63%)	0.22	Gehring et al. 2011
AT3G28980	3' (+2006bp until +2096bp)	0.9998	59/90 (65%)	0.07	Gehring et al. 2011
AT4G29640	3' (+769bp until +847bp)	0.9998	53/78 (67%)	0.19	Wolff et al. 2011
AT4G29650	5' (-953bp until-875bp)	0.9998	53/78 (67%)	0.19	Wolff et al. 2011

This table lists the imprinted loci in *Arabidopsis*, which have a *cis*-element 3kb up- or downstream of the respective ORF that shows sequence similarity with the *MEA*-ICR. The location is displayed in base pairs upstream of the start codon (3') or downstream of the stop codon (5') of the imprinted gene. The p-value is the smallest sum probability calculated by the WU-BLAST tool (www.arabidopsis.org). The identities are common base pairs per length of the conserved element between the *MEA*-ICR and the respective *cis*-element. The rows in bold indicate the *cis*-elements with the highest similarity that were used for the motif analysis shown in Table S4.

Table S4. Common motifs identified in the *MEA*-ICR and the 6 ICR-like sequences analyzed

Motif name	Motif sequence	PLACE number	occurrence in *MEA*-ICR
GT-1 binding site	GRWAAW	S000198	9
DOF binding element	AAAG	S000265	5
Pollen element; lat52	AGAAA	S000245	4
TATABOX5	TTATTT	S000203	2
GATABOX	GATA	S000039	1
polyA signal box	AATAAT	S000088	1

The table shows an overview of the motifs identified in the *MEA*-ICR and the six most similar *cis*-elements up- or downstream of other imprinted loci (i.e. the ones with the smallest p-values, see Supplemental Table S3). Motif analysis was performed using the PLACE database (Higo et al. 1999).

Table S5. Quantification of *MEA* and *mea* transcript levels

Cross	*MEA* x *mea*			*MEA;met1-3* x *mea*			
	MEA^m transcripts			MEA^m transcripts			
	Average	SEM	Variance	Average	SEM	Variance	p-value
BF	1.1148	0.1132	0.0192	1.5821	0.3078	0.1421	0.0464
1DAP	0.8008	0.4749	0.3383	0.7385	0.0506	0.0038	0.8071
2DAP	0.3234	0.0401	0.0024	0.3119	0.0372	0.0021	0.6964
3DAP	0.3399	0.3126	0.1465	0.1417	0.0852	0.0109	0.2882
4DAP	0.1148	0.0107	0.0002	0.1098	0.0580	0.0051	0.8730
	mea^p transcripts			mea^p transcripts			
	Average	SEM	Variance	Average	SEM	Variance	p-value
1DAP	0.0259	0.0167	0.0004	0.0244	0.0201	0.0006	0.9163
2DAP	0.0079	0.0052	0.0000	0.112	0.0053	0.0000	0.4295
3DAP	0.0106	0.0072	0.0001	0.0166	0.0130	0.0003	0.4631
4DAP	0.0079	0.0030	0.0000	0.0108	0.0080	0.0001	0.5440

Cross	*mea* x *MEA*			*mea* x *MEA;met1-3*			
	mea^m transcripts			mea^m transcripts			
	Average	SEM	Variance	Average	SEM	Variance	p-value
BF	13.6641	6.5242	63.8474	17.8910	14.2141	303.0629	0.6175
1DAP	8.6833	0.9223	1.2761	14.4017	8.8134	116.5128	0.2664
2DAP	1.6167	1.2343	2.2854	2.1648	1.3488	2.7289	0.5811
3DAP	0.4402	0.2797	0.1173	0.4523	0.2878	0.1243	0.9549
4DAP	0.4313	0.2400	0.0864	0.4975	0.3763	0.2124	0.7817
	MEA^p transcripts			MEA^p transcripts			
	Average	SEM	Variance	Average	SEM	Variance	p-value
1DAP	0.1563	0.0442	00029	0.2000	0.0750	0.0084	0.3722
2DAP	0.2515	0.0412	0.0025	0.2800	0.1046	0.0164	0.6387
3DAP	0.1583	0.0746	0.0083	0.0600	0.0196	0.0006	0.0635
4DAP	0.1752	0.0915	0.0126	0.1700	0.1205	0.0218	0.09484

The Table shows the quantification of the *MEA* and *mea* (*mea-2*) transcript levels in the four different crosses described in **Fig. 4**. Transcript levels were normalized to *Actin11* and expressed relative to one reference *MEA* wild-type sample. The p-value designates the level of significance relative to the difference between MEA^m transcripts in each two types of crosses and he level of significance relative to the difference of MEA^p transcripts in each two types of crosses. Underlined values highlight the fact that the level of de-repressed paternal transcripts in maternal *mea* mutants represents only 19.5% (0.1563/0.8008) of the amount of maternal transcripts in the maternal wild-type background and represent only 1.8% (0.1563/8.6833) of the amount of maternal transcripts in maternal *mea* mutants.

Average, average of 3 independent $2^{DDC}{}_T$ values, SEM, standard error of the means corresponding to 3 independent biological replicates for each cross and each time point, BF, before fertilization, DAP, days after pollination.

Table S6. Primer sequences

Primer Name	Sequence 5' to 3'
Act11F	AACTTTCAACACTCCTGCCATG
Act11R	CTGCAAGGTCCAAACGCAGA
CK176	AATTCCATGGTAACCACTCGCCTCTTC
DP200	GTTGAATTCATCGCCCAAGCTTGTGC
DP202	TTGGAATTCGAGCTGTCAAACGTCAAGC
DP248	AATTCCATGGAGAAGGTTAGTTTCACTCC
DP71	ACCGAGCTCCATGGTCGACAAAAGTGATTTGTTGTC
Ds3.1	CGATTACCGTATTTATCCCGTTCG
Ds5.1	CCGTTTACCGTTTTGTATATCCCG
FWA-Bregion-F	TGGTTGTTTAAGGTTGYTTTTAGYAYA
FWA-Bregion-R	ACAACAAAAATCTRATTRTCARTATCCT
HW25	TACACTACCAACTGATTACG
HW26	TCTCCTCTCATTGGACATGC
HW35	GGTAGTACTAAGCAAGTCCAAATACG
HW45	GCTTGTGTCGCCTTCGGA
HW47	CGTTTAACCTCTGCAACCACC
MEA2retiAS1	CACCAAGAGTGCCATCTCCA
MEA2retiS1	GGATTGCAACAATCGCTTTG
MEA2retiS2	CCAATGCACAAATCGACAATG
MEA5'10	CGGGATCCCCATTGAACATTAATTTAAGTC
MEA5'11	CGGGATCCAAGCTTAAATAGAAAACTAGC
MEA5'14	GTTGAATTCCCAAAATGTATATATTGATCTATCAAC
MEA5'15	GTTGAATTCAAGCAAGTCCAAATACGTTTCTTCC
MEA5'8	CGGGATCCGAGAGCCGTTGTGGCAGTGACC
MEA5'9	CGGGATCCCTTTGAGCATTAAGAGGTCG
MEAretirev2	GGTAGGAAGAACCAATCCGATCT
MEAS11	TCTGATGTTCATGGATGGGG
MEDEA-A5-F	ATTTTTTTATTGGTTTATTAAATTAAGTTGTT
MEDEA-A5-R	ACAAAAAAATTCAACCCTAAATATTATTA
MEF1	GATTGTGTCTCTACTACAGAGGC
MEF2	GCCTGGTCAAGTGGACTTCATC
MER2	CCATTCTTCACAGAGCATGCC
TL2	TGGACGTGAATGTAGACACGTCG
UA171	GCTTCACACCATCAATCGTTTGAC
UA8	CGAGCTCCTAACGAGCTGGACGG

The Table shows the summary of all primers used in this work. All sequences are given in the 5' to 3' direction.

Appendix A4: Tables of all candidates isolated in the forward genetics screen

Table A4.1. Primary maternal activator candidates of screen #1.

	M1		M2			BC1			BC2			
line	GUS+ seeds	fertility	GUS	fertility	individuals	GUS	fertility	individuals	GUS	fertility	individuals	General comments
1-4V	<50%	n.d.	wt	n.d.								wild type only in M2
1-10V	<50%	n.d.	m	50% i.O + 25% a.s.	1 of 24	wt	wt	0 of 40				wild type only in BC1
1-17H	<50%	n.d.	m	50% i.O.	1 of 12	m	m	8 of 18	m	m	6 of 17	**matA1**; infertile ovules only
3-19H	<50%	n.d.	m	>50% i.O.	1 of 13	?	?	?				strong hybrid vigour in BC1
4-5V	<50%	n.d.	wt	n.d.								wild type only in M2
5-1H	<50%	n.d.	wt	n.d.								wild type only in M2
6-3V	<50%	n.d.	m	50% a.s. (*mea*-like)	2 of 12	wt	m	0 of 48				unlinked s.a. phenotype
7-22V	<50%	n.d.	wt	n.d.								wild type only in M2
8-11H	<50%	n.d.	wt	n.d.								wild type only in M2
10-13V	<50%	n.d.	wt	n.d.								wild type only in M2
10-16V	<50%	n.d.	wt	n.d.								wild type only in M2
11-20V	<50%	n.d.	wt	n.d.								wild type only in M2
12-21H	<50%	n.d.	wt	n.d.								wild type only in M2
13-20V	<50%	n.d.	wt	n.d.								wild type only in M2
14-23V	<50%	n.d.	wt	n.d.								wild type only in M2
15-6H	<50%	n.d.	wt	n.d.								wild type only in M2
15-11H	<50%	n.d.	wt	n.d.								wild type only in M2
16-19H	<50%	n.d.	wt	n.d.								wild type only in M2

Table A4.2. Primary paternal repressor candidates of screen #1.

	M1		M2		BC1		BC2		BC3		BC4		BC5		
line	GUS+ seeds	individuals	GUS	individuals	GUS	individuals	GUS	individuals	GUS	individuals	GUS	individuals	GUS	individuals	General comments
2-20H	50%[1]	1 of 1	30%[1]	4 of 12	30-50%[2]	9 of 14	15-25%[3]	5 of 10	10-50%[3]	24 of 24?	n.d.	n.d.			**patR2**
8-13V	40%[1]	1 of 1	20-40%[1]	3 of 11	30%[2]	11 of 35	n.d.	n.d.							**patR8**
9-8H	60%[1]	1 of 1	20-50%[1]	4 of 12	40-50%[2]	22 of 44	10-15%[3]	7 of 15	12-56%[3]	22 of 22?	10-44J%	20 of 20?	10-25%[2]	9 of 24!	**patR9**
1-15V	30%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
1-15H	30%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
2-18V	30%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
3-7H	40%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
3-18H	20%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
5-2V	30%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
5-4V	35%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
7-5V	30%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
9-14H	60%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
9-23H	60%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
11-2V	50%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
11-4V	50%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
11-14H	50%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up
16-28H	20%	1 of 1	n.d.	n.d.	n.d.	n.d.									not followed up

[1]: on *dde2-3* (Ler) mothers

[2]: on *dde2-2* (Col-0) mothers (von Malek et al. 2002)

[3]: on emasculated FUK (*dme-4/DME* wild-type segregant) mothers

Table A4.3. Primary maternal activator candidates of screen #2 displaying infertile ovules. Only the M1 generation was analyzed due to the high probability of being a false positive candidate.

line	GUS+ seeds	plump seeds	aborted seeds	fertility	comments
20-27V	30	15	0	i.O.	2 intensities
18-26H	35	34	0	i.O.	2 intensities
26-18H	1	0	0	i.O.	sterile
5-10V	2	0		i.O.	sterile
9-8H	7	3	3	i.O.	sterile
25-11H	12	6	2	i.O.	sterile
23-5V	18	6	0	i.O.	sterile
1-3H	23	0		i.O.	sterile
9-22V	7	8	4	i.O.	sterile?
26-3V	8	4	2	i.O.	sterile?
6-19V	25	7		i.O.	sterile?
23-20V	5	20	0	i.O.	
27-10V	6	24	1	i.O.	
19-12H	7	20	0	i.O.	
6-8V	8	30		i.O.	
30-12V	9	20	0	i.O.	
26-28H	10	13	1	i.O.	
5-2V	11	19		i.O.	
22-22H	11	24	0	i.O.	
26-13V	11	15	0	i.O.	
28-25V	12	23	0	i.O.	
24-24H	13	26	0	i.O.	
28-4V	13	23	0	i.O.	
29-4H	13	18	1	i.O.	
3-2H	14	17		i.O.	
8-15V	14	24	0	i.O.	
23-24V	14	25	2	i.O.	
29-22V	14	18	0	i.O.	
6-11H	15	24		i.O.	
9-10H	15	24	0	i.O.	
13-25V	15	29	0	i.O.	
19-7V	15	22	1	i.O.	
12-2H	16	13	0	i.O.	
15-17V	16	23	0	i.O.	
20-13H	16	27	0	i.O.	
25-9V	16	17	0	i.O.	
26-23H	16	12	2	i.O.	
2-17V	17	39		i.O.	
3-12H	17	20		i.O.	
7-9H	17	19	0	i.O.	
30-18H	17	29	0	i.O.	
6-23H	18	13		i.O.	
23-3V	18	28	0	i.O.	
23-14V	18	31	0	i.O.	
1-26V	19	8		i.O.	
1-28V	19	36		i.O.	
7-28H	19	18	0	i.O.	

line	GUS+ seeds	plump seeds	aborted seeds	fertility	comments
18-22V	19	21	0	i.O.	
1-16H	20	15		i.O.	
5-7H	20	26		i.O.	
6-15V	20	19		i.O.	
10-12V	20	19	2	i.O.	
30-8V	20	21	0	i.O.	
30-27V	20	16	1	i.O.	
1-3V	21	25		i.O.	
4-8V	21	22		i.O.	
7-17V	21	21	0	i.O.	
8-17H	21	16	0	i.O.	
14-20H	21	21	1	i.O.	
16-11V	21	25	0	i.O.	
18-7V	21	26	0	i.O.	
6-3V	22	14		i.O.	
12-3V	22	24	0	i.O.	
12-19H	22	17	3	i.O.	
16-20H	22	14	0	i.O.	
20-19H	22	22	3	i.O.	
30-6V	22	16	0	i.O.	
9-14V	23	18	1	i.O.	
11-22V	23	32	0	i.O.	
23-26V	23	25	0	i.O.	
17-5H	24	31	0	i.O.	
24-17V	24	26	0	i.O.	
28-2V	24	20	3	i.O.	
1-11V	25	16		i.O.	
5-28H	25	21		i.O.	
8-10V	25	30	0	i.O.	
25-5V	25	11	0	i.O.	
5-8V	26	34		i.O.	
9-19V	27	25	1	i.O.	
10-8H	27	19	0	i.O.	
11-6H	27	24	0	i.O.	
13-27H	27	14	0	i.O.	
19-13V	27	15	2	i.O.	
24-16H	27	18	0	i.O.	
29-26V	27	25	0	i.O.	
6-3H	28	26		i.O.	
11-19V	28	28	0	i.O.	
19-10V	28	28	0	i.O.	
20-4V	28	23	0	i.O.	
8-11H	29	19	0	i.O.	
14-23H	29	24	3	i.O.	
3-16H	30	41		i.O.	
3-20H	30	43		i.O.	
10-28V	30	28	2	i.O.	
4-16H	31	36		i.O.	
18-20H	32	25	0	i.O.	
4-14H	34	31		i.O.	

Table A4.4. Primary maternal activator candidates of screen #2 displaying infertile ovules and seed abortion. Only the M1 generation was analyzed due to the high probability of being a false positive candidate.

line	GUS+ seeds	plump seeds	aborted seeds	fertility	comments
25-16V	24	16	5	s.a. and i.O.	2 intensities
15-14H	31	18	8	s.a. and i.O.	2 intensities
24-2H	31	26	6	s.a. and i.O.	2 intensities
29-11H	31	22	3	s.a. and i.O.	2 intensities
15-21V	33	27	12	s.a. and i.O.	2 intensities
13-2V	34	28	4	s.a. and i.O.	2 intensities
11-8V	38	28	10	s.a. and i.O.	2 intensities
25-2V	38	26	8	s.a. and i.O.	2 intensities
16-7H	33	16	14	s.a. and i.O.	2 stages s.a.
29-21V	0	0	19	s.a. and i.O.	sterile
28-11H	5	0	6	s.a. and i.O.	sterile
28-22V	14	0	22	s.a. and i.O.	sterile
19-15V	25	1	7	s.a. and i.O.	sterile
26-4H	8	4	6	s.a. and i.O.	sterile?
16-13V	20	9	3	s.a. and i.O.	sterile?
22-15H	3	12	12	s.a. and i.O.	
26-2V	5	12	3	s.a. and i.O.	
18-5H	6	11	3	s.a. and i.O.	
25-4V	6	12	6	s.a. and i.O.	
30-1V	6	8	5	s.a. and i.O.	
29-15V	8	15	9	s.a. and i.O.	
27-9H	9	37	1	s.a. and i.O.	
28-1V	9	25	7	s.a. and i.O.	
28-16H	9	13	15	s.a. and i.O.	
19-3V	11	18	4	s.a. and i.O.	
22-28V	11	19	6	s.a. and i.O.	
28-24V	11	22	6	s.a. and i.O.	
26-17V	12	9	9	s.a. and i.O.	
27-18H	14	6	7	s.a. and i.O.	
7-3V	15	7	15	s.a. and i.O.	
9-4V	15	15	4	s.a. and i.O.	
29-10V	15	16	8	s.a. and i.O.	
14-22V	16	6	14	s.a. and i.O.	
18-25V	16	8	2	s.a. and i.O.	
24-20V	16	16	7	s.a. and i.O.	
24-21V	16	19	3	s.a. and i.O.	
24-22H	16	15	6	s.a. and i.O.	
28-8H	16	16	4	s.a. and i.O.	
28-12V	16	15	4	s.a. and i.O.	
30-16H	16	13	6	s.a. and i.O.	
12-8H	17	9	10	s.a. and i.O.	
16-1V	17	14	4	s.a. and i.O.	
17-26H	17	28	2	s.a. and i.O.	
28-14H	18	7	10	s.a. and i.O.	
29-13V	18	22	5	s.a. and i.O.	
13-7H	19	15	9	s.a. and i.O.	
24-23V	19	17	4	s.a. and i.O.	
26-20H	19	26	4	s.a. and i.O.	
27-12H	19	23	7	s.a. and i.O.	
9-27V	20	14	8	s.a. and i.O.	
14-14V	20	15	8	s.a. and i.O.	
22-24V	20	22	4	s.a. and i.O.	
29-17V	20	15	6	s.a. and i.O.	
30-16V	20	15	4	s.a. and i.O.	
7-18V	21	8	27	s.a. and i.O.	
15-20V	21	16	4	s.a. and i.O.	
27-23V	21	22	9	s.a. and i.O.	
1-27H	22	35		s.a. and i.O.	
7-13H	22	10	6	s.a. and i.O.	
16-27V	22	24	7	s.a. and i.O.	
22-25H	22	31	2	s.a. and i.O.	
23-27H	22	14	14	s.a. and i.O.	
24-5H	22	13	15	s.a. and i.O.	
26-14V	22	9	2	s.a. and i.O.	
26-18V	22	16	13	s.a. and i.O.	
28-13H	22	20	11	s.a. and i.O.	
30-14H	22	22	10	s.a. and i.O.	
22-19V	23	27	10	s.a. and i.O.	
23-2V	23	18	5	s.a. and i.O.	
26-24H	23	25	7	s.a. and i.O.	
29-16H	23	17	16	s.a. and i.O.	
30-7V	23	4	10	s.a. and i.O.	
17-11V	24	8	10	s.a. and i.O.	
30-1H	24	15	4	s.a. and i.O.	
12-10H	25	37	2	s.a. and i.O.	
14-24H	25	13	6	s.a. and i.O.	
20-24H	25	9	8	s.a. and i.O.	
13-2H	26	23	9	s.a. and i.O.	
25-25V	26	23	9	s.a. and i.O.	
27-24H	26	20	5	s.a. and i.O.	
16-27H	27	11	6	s.a. and i.O.	
22-21V	27	27	5	s.a. and i.O.	
16-23H	29	31	2	s.a. and i.O.	
19-1H	29	28	9	s.a. and i.O.	
19-16V	31	31	4	s.a. and i.O.	

Table A4.5. Primary maternal activator candidates of screen #2 displaying seed abortion.

	M1							BC1/M2										
line	GUS+ seeds	plump seeds	aborted seeds	fertility	stage of s.a.	comments	BC1?	MR	line	GUS wt	GUS mut	n=	sa. wt	sa. m	n=	linked sa?	GUS	sa
6-5H	36	32		s.a.	?	2 intensities	no	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
15-3V	30	29	15	s.a.	ws	2 intensities	yes	194	BC1 15-3V	23	1	24	2	1	3	yes	m	iO
16-23V	33	24	13	s.a.	ws	2 intensities	yes	198	BC1 16-23V	24	0	24	1	5	6	no	wt	ws&iO
18-12V	32	34	10	s.a.	ws	2 intensities	yes	203	BC1 18-12V	24	0	24	4	2	6	no	wt	ws
27-15H	28	22	18	s.a.	ws & bs	2 intensities	yes	219	BC1 27-15H	24	0	24	1	5	6	no	wt	ws&bs&iO
14-8V	32	28	18	s.a.	ws & bs	2 intensities	yes	192	BC1 14-8V	21	3?	21	1	5	6	no	wt?	ws&bs
13-24H	32	26	15	s.a.	ws & bs	2 intensities	yes	189	BC1 13-24H	24	0	24	2	0	2	no	wt	wt
17-23H	37	25	15	s.a.	ws & bs	2 intensities	yes	202	BC1 17-23H	24	0	24	4	2	6	no	wt	ws
12-27V	31	32	13	s.a.	ws & bs	2 intensities	yes	187	BC1 12-27V	24	0	24	1	3	4	no	wt	ws
19-9H	26	24	10	s.a.	?	few i.O.	yes	205	BC1 19-9H	21	3	24	7	5	12	no	m	ws&iO
27-26V	7	0	38	s.a.	ws	sterile?	yes	214	BC1 27-26V	20	4	24	8	16	24	no	m	ws&iO
4-7H	30	35		s.a.	?		yes	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
24-7H	15	19	14	s.a.	bs		no	236	M2 24-7V	10	2	12	4	8	12	no	m	wt
24-13V	26	28	13	s.a.	bs		no	238	M2 24-13V	10	2	12	9	3	12	**yes, but..**	**m**	**ws&iO**
25-8V	20	35	7	s.a.	bs		no	240	M2 25-8V	11	0	11	2	1	3	no	wt	iO
24-11H	28	32	6	s.a.	bs		no	237	M2 24-11H	10	1	11	3	3	6	no	m	iO
15-15V	22	28	20	s.a.	ws		yes	195	BC1 15-15V	22	2	24	22	2	24	**yes**	**m**	**ws**
28-3V	10	19	16	s.a.	ws		yes	220	BC1 28-3V	24	0	24	4	2	6	no	wt	ws
30-4V	11	22	14	s.a.	ws		yes	216	BC1 30-4V	24	0	24	2	4	6	no	wt	bs&iO
28-18V	18	22	14	s.a.	ws		yes	215	BC1 28-18V	14	10	24	4	20	24	no	m	ws&bs
26-1V	22	27	14	s.a.	ws		yes	209	BC1 26-1V	24	0	24	n.d.	n.d.	0	no	wt	?
13-28V	22	38	12	s.a.	ws		yes	190	BC1 13-28V	24	0	24	15	9	24	no	wt	ws
19-17H	29	26	12	s.a.	ws		yes	207	BC1 19-17H	24	0	24	n.d.	n.d.	0	no	wt	?
17-9V	25	24	11	s.a.	ws		yes	200	BC1 17-9V	24	0	24	6	0	6	no	wt	wt
12-13H	26	38	11	s.a.	ws		no	227	M2 12-13H	12	0	12	12	0	12	no	wt	wt
13-12H	18	30	10	s.a.	ws		yes	188	BC1 13-12H	24	0	24		2	2	no	wt	ws
26-17H	19	22	10	s.a.	ws		yes	218	BC1 26-17H	24	0	24	3	3	6	no	wt	ws
22-21H	23	28	10	s.a.	ws		no	232	M2 22-21H	9	3	12	2	10	12	no	m	ws&bs&iO
26-24V	13	30	8	s.a.	ws		yes	212	BC1 26-24V	24	0	24	n.d.	n.d.	0	no	wt	?
22-19H	16	31	8	s.a.	ws		no	231	M2 22-19H	10	2	12	1	5	6	no	m	iO
23-19H	22	35	8	s.a.	ws		no	235	M2 23 19H	11	1	12	2	4	6	no	m	iO
16-18V	24	35	6	s.a.	ws		yes	197	BC1 16-18V	23	1	24	24	0	24	no	wt	wt
22-8H	24	40	6	s.a.	ws		no	228	M2 22-8H	11	1	12	11	1	12	yes	m	iO
22-9H	19	23	19	s.a.	ws & bs		no	229	M2 22-9H	7	5	12	2	10	12	no	m	ws&iO
12-7H	14	20	18	s.a.	ws & bs		yes	186	BC1 12-7H	24	0	24	2	0	2	no	wt	wt
18-23V	19	25	12	s.a.	ws & bs		yes	204	BC1 18-23V	21	3	24	9	3	12	**yes**	**m**	**ws**
23-7V	24	29	11	s.a.	ws & bs		no	233	M2 23-7V	12	0	12	0	4	4	no	wt	ws
11-20H	17	31	9	s.a.	ws & bs		yes	183	BC1 11-20H	24	0	24	6	0	6	no	wt	wt
16-8H	19	34	8	s.a.	ws & bs		yes	196	BC1 16-8H	5	18	23	5	18	23	**yes**	**m**	**iO**
11-4H	26	43	3	s.a.	ws & bs		yes	182	BC1 11-4H	24	0	24	1	5	6	no	wt	iO

> matA15

> matA18

Table A4.6. Primary maternal activator candidates of screen #2 with full fertility.

	M1						BC1/M2										
line	GUS+ seeds	plump seeds	aborted seeds	fertility	comments	BC1?	MR	line	GUS wt	GUS mut	n=	sa. wt	sa. m	n=	linked sa?	GUS	sa	
28-6H	13	31	0	s.a.??	evtl. s.a.	yes	221	BC 28-6H	24	0	24	6	0	6	no	wt	wt	
25-21V	9	42	1	s.a.??		no	241	M2 25-21V	11	1	12	5	7	12	**yes but..**	**m**	**ws&iO**	**> matA25**
28-9H	13	33	0	s.a.??		yes	222	BC1 28-9H	21	3	24	21	3	24	**yes**	**m**	**iO**	
23-12H	14	38	0	s.a.??		no	234	M2 23-12H	12	0	12	2	2	4	no	wt	iO	B-class mutant!
24-26H	15	46	0	s.a.??		no	239	M2 24-26H	9	3	12	9	3	12	**yes**	**m**	**iO**	
11-24V	16	39	?	s.a.??		yes	184	BC1 11-24V	16	0	16	6	0	6	no	wt	wt	
25-28V	16	45	0	s.a.??		no	242	M2 25-28V	11	1	12	1	5	6	no	m	iO	
6-18V	20	48		s.a.??		no	225	M2 6-18V	12	0	12	3	0	3	no	wt	wt	
14-6V	20	44	1	s.a.??		yes	191	BC1 14-6V	24	0	24	2	0	2	no	wt	wt	
14-15V	20	46	0	s.a.??		yes	193	BC1 14-15V	24	0	24	2	0	2	no	wt	wt	
17-8V	21	32	1	s.a.??		yes	199	BC1 17-8V	24	0	24	6	0	6	no	wt	wt	
22-17V	21	42	0	s.a.??		no	230	M2 22-17H	11	1	12	5	1	6	**yes**	**m**	**iO**	
10-2H	23	50	0	s.a.??		no	226	M2 10-2H	9	3	12	9	3	12	**yes**	**m**	**iO**	
20-2H	23	31	1	s.a.??		yes	208	BC1 20-2H	24	0	24	n.d.	n.d.	0	no	wt	?	
11-26H	24	44	?	s.a.??		yes	185	BC1 11-26H	23	0	23	6	0	6	no	wt	wt	
17-17V	24	38	0	s.a.??		yes	201	BC1 17-17V	23	1	24	5	1	6	**yes**	**m**	**iO**	
26-11H	24	40	0	s.a.??		yes	217	BC1 26-11H	24	0	24	6	0	6	no	wt	wt	

Table A4.7. Primary maternal activator candidates of screen #2 that show ectopic reporter expression in the sporophytic seed coat. Only the M1 generation was analyzed and the candidates were not followed up.

	M1			
line	**GUS+ seeds**	**plump seeds**	**aborted seeds**	**comments**
15-4V	45	39	0	sporophytic staining
10-13H	48	51	0	sporophytic staining

Table A4.8. Maternal Activator candidate 18 (matA18); segregation and transmission efficiency.

line	generation	mutant	wild type	% mutant
MR255	BC2	49	159	23.56
MR264	BC3	13	35	27.08
MR272	BC4	14	20	41.18
MR272	BC5	10	38	20.83
	TOTAL:	**86**	**252**	**25.44**

line	generation	mutant	wild type	% mutant
MR265	OC2	44	48	47.83

male transmission	0.51	n=338	all backcrosses
female transmission	0.73	n=96	MR272.7 x MR151

Table A4.9. Maternal Activator candidate 25 (matA25); segregation and transmission efficiency.

line	generation	mutant	wild type	% mutant
MR270	BC2	69	215	24.30
MR278	BC3	2	36	5.26
	TOTAL:	**71**	**251**	**22.05**

line	generation	mutant	wild type	% mutant
MR289	OC2	36	90	28.57

male transmission	0.44	n=322	all backcrosses
female transmission	n.d.	n.d.	

GENETICS

Supporting Information
http://www.genetics.org/content/suppl/2012/05/26/genetics.112.141341.DC1

SNP-Ratio Mapping (SRM): Identifying Lethal Alleles and Mutations in Complex Genetic Backgrounds by Next-Generation Sequencing

Heike Lindner, Michael T. Raissig, Christian Sailer, Hiroko Shimosato-Asano, Rémy Bruggmann, and Ueli Grossniklaus

DOI: 10.1534/genetics.112.141341

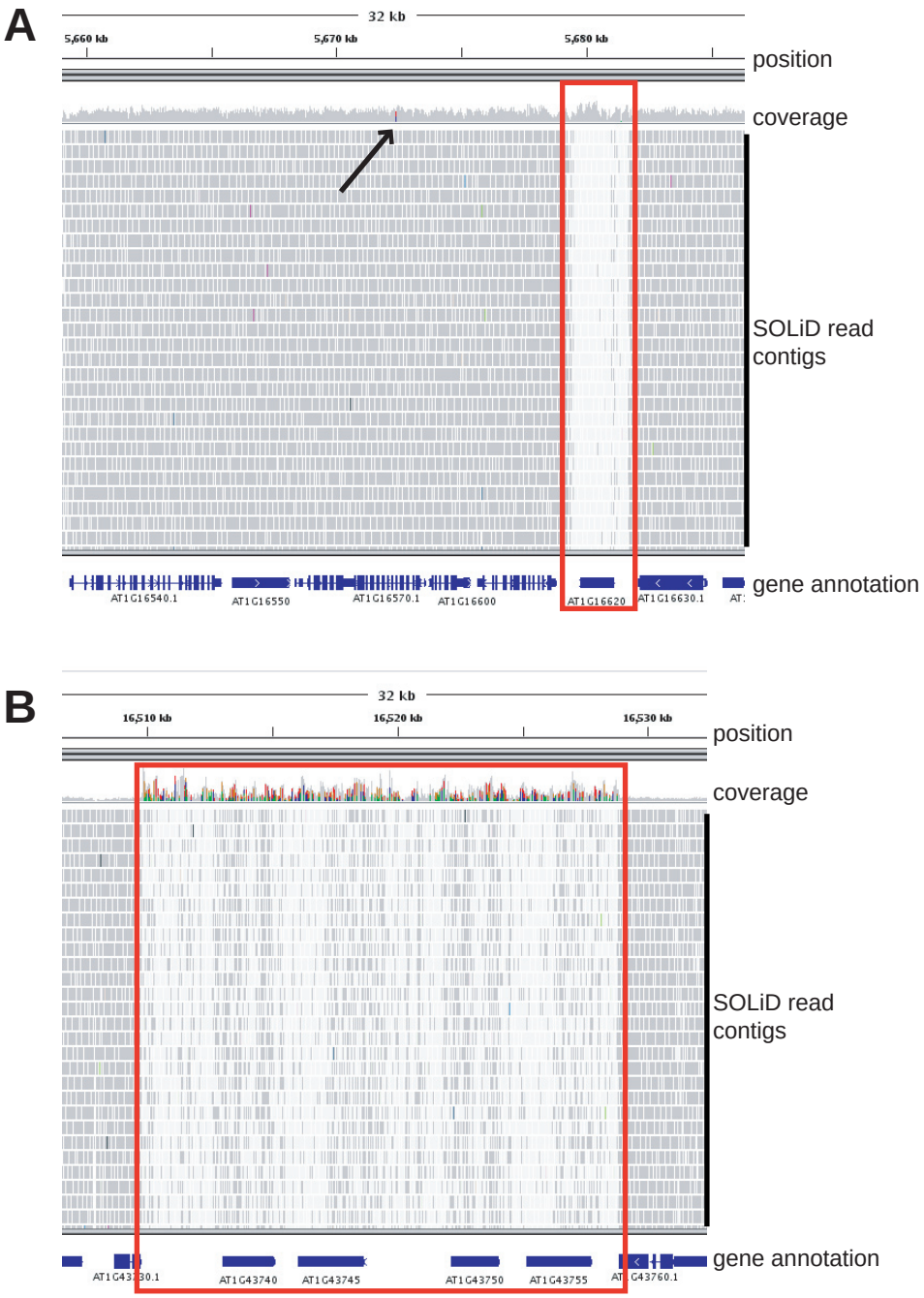


Figure S1 Integrative Genomics Viewer (www.broadinstitute.org/igv/) screenshots of mapped reads around the causative SNP in *At1g16570* (A) and in a transposable element gene region on chromosome I (B). (A) The causative SNP (arrow) is in a region with high qualitiy reads (dark grey colored contigs) but the transposable element gene *At1g16620* is only covered by low mapping quality reads (light grey colored contigs, red box). (B) The transposable element genes *At1g43740* until *At1g43755* (red box) are mainly covered by low quality reads (light grey contigs). As a consequence, the SNP density is very high, indicated as colored bars in the coverage plot on top.

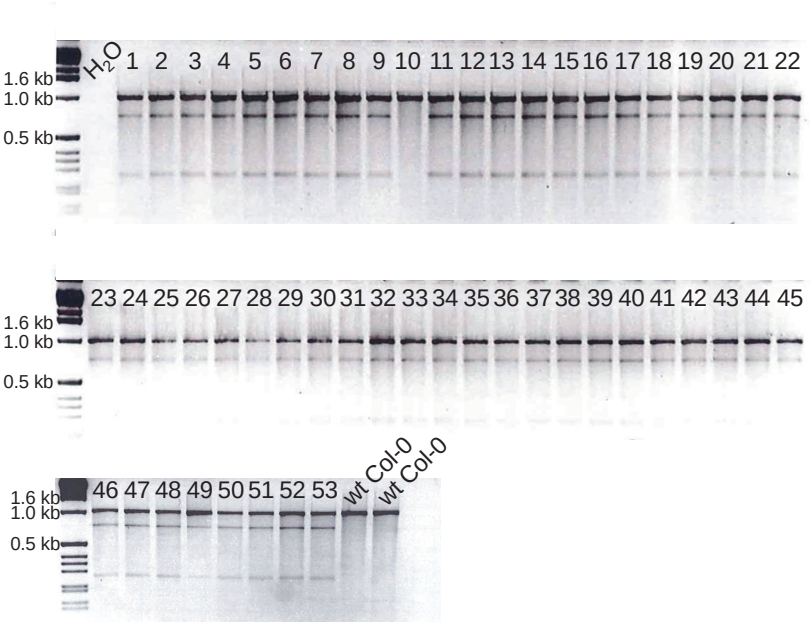


Figure S2 SURVEYOR™ nuclease digest of sequenced samples. The SNP region of the *At1g16570* gene was amplified from each of the 53 DNA samples that had been pooled for sequencing, two Col-0 controls and a H_2O control. PCR products were digested with the SURVEYOR™ nuclease cleaving single base pair mismatches in heteroduplex DNA (TILL et al. 2004). The undigested wild-type band is 989 bp, whereas any sample containing the SNP displays an undigested band at 989 bp and the digestion products at 719 bp and 217 bp. Individual 10, showing a wild type band only, was a sampling mistake (see text).

File S1

Supporting Methods

Plant material:

0.23g Col-0 seeds with the *Lat52::GFP* reporter were treated with 0.15% ethane methyl sulfonate (EMS) for 8h. M1 plants were grown as previously described (HUCK et al. 2003). For screening, siliques of 1800 M1 plants were collected 2 days after pollination and prepared for aniline blue staining of callose in pollen tubes (HUCK et al. 2003). Images were captured with a Leica DM6000B epifluorescence microscope. One line with high penetrance of the pollen tube overgrowth phenotype was backcrossed twice to the Col-0 wild-type accession, using the wild-type line (Col-0) as pollen donor.

Preparation of genomic DNA and SOLiD sequencing:

Genomic DNA was extracted with a QIAGEN DNeasy Plant Mini Kit from 53 BC2 individuals displaying the mutant phenotype. By mixing DNA in equimolar concentrations, a fragment library was produced using the SOLiD Fragment Library Construction kit (Applied Biosystems). 1 µg of genomic DNA was fragmented to a size range of 100–150 bp with the use of a Covaris S2 (Covaris) and analyzed on an Agilent Bioanalyzer 2100 DNA Chip 7500. Fragments were end-repaired using polishing enzymes and ligated with barcoded adapters in the presence of T4 DNA Ligase. The ligated DNA was size selected to approximately 200bp using AMPure beads XP (Agencourt). This product was nick translated and amplified by PCR with primers complementary to the sequence of the SOLiD 4 adaptors. After quality control on an Agilent Bioanalyzer DNA Chip 1000 the libraries were pooled and used for e-PCR based on a concentration of 0.5pM. Approximately one half of a SOLiD 4 sequencing slide (Applied Biosystems) was used.

Sequence data analysis:

The paired end (50x35bp) sequence reads from SOLiD platform were mapped to the *Arabidopsis thaliana* genome assembly TAIR version 10 (ftp.arabidopsis.org/home/tair/Sequences/whole_chromosomes/) using Bioscope version 1.3.1 (default settings) from Applied Biosystems (Life Technologies, Carlsbad, CA, USA). After mapping, the SNPs were called by using the DiBayes algorithm (Life Technologies, Carlsbad, CA, USA) with high and medium stringency settings. The obtained SNPs were classified into intronic, intergenic, coding (synonymous and non-synonymous amino acid substitutions) and splice-site variants with a custom software pipeline.

Statistical identification of causative SNP candidates

The results of the sequence data analysis were tabulated in the following structure: chromosome, position on the chromosome in bp, number of reference counts (reads), number of variant counts, zygosity (either “homozygous” or “heterozygous”). The table in this structure was loaded into R as tab-delimitated text file. To identify the 1:1 segregating

SNPs the dataset was filtered in 3 steps: First, all homozygous reads were removed. Second, all reads with a coverage less than 50-fold were removed (see Table S1). Third, F~Binomial (H_0: Π = 0.5; H_A: Π ≠ 0.5) was calculated for each SNP. All SNPs with $\alpha < 0.05$ and $\alpha > 0.95$ were removed. This resulted in a list of 1:1 segregating SNPs (n=118).

SNPs neighboring the causative SNP are genetically linked and thus follow an expected pattern of co-segregation. The expected pattern of co-segregation of two neighboring SNPs on each side of each 1:1 segregating SNP was calculated. This was done with the complete dataset of heterozygous SNP without filtering for low and high coverage in 3 steps:

- First, the genetic distance (in cM) between the SNPs was calculated by dividing the physical distance (bp) between the SNPs by the mean physical distance per cM (357,042 bp/cM). The mean genetic distance was calculated using the physical map and the most recent, sex-specific genetic map of *Arabidopsis thaliana* (TAIR version 10 (www.arabidopsis.org); GIRAUT et al. 2011). In *Arabidopsis*, dramatic differences between male and female meiosis have been found (genetic map length: male=575cM, female=332cM) and since we crossed the mutants always as females, the female genetic map length was used here. However, we obtained the same results using the overall mean genetic distance of 1cM per 198,556 bp (i.e. combined male and female recombination rates: genetic map length=597cM; LISTER AND DEAN, 1993). Since the mean physical distance in bp per cM can be estimated for most organisms used in genetic research, this procedure is generally applicable. If more accurate combined genetic and physical maps are available, local recombination rates could be integrated into this procedure.
- Second, the expected number of variant reads for a 1:1 class SNP was calculated by dividing the number of total reads (coverage) per SNP by two (diploid organism).
- Third, the expected number of variant reads for each neighboring and co-segregating SNP was calculated according to the genetic distance between the SNPs based on the mean physical distance per cM. This was done for each of the two neighboring SNPs on each side of the 1:1 segregating SNP.

To test which SNPs differ from the expected pattern of co-segregation, a χ^2 goodness of fit test was applied. The calculated expected number of variant reads was subtracted from the observed number of variant reads and this value was squared, followed by division with the calculated expected number of variant reads. This was done for each of the 2 neighboring SNPs on each side of the 1:1 segregating SNP, giving 5 χ^2-values, which were summed up (for $\alpha < 0.05$ and df = 4, $\chi^2 = 9.49$; en.wikipedia.org/wiki/Chi-squared_distribution). All SNPs with a $\chi^2 > 9.49$ were discarded, because they differed significantly from the genetic expectations.

The R-script is available as an extra file (File S2) and can be downloaded at http:/

Confirmation of mutation:

The SNP region of the *At1g16570* gene was amplified from each of the 53 DNA samples, that had been pooled for sequencing, two Col-0 controls and a H_2O control with the primers 5`-CTTTTGATCTTGCAGCCACA-3`and 5`-

ACGTGCAGCGTTTGTTGTAG-3 and ExTaq Polymerase (TaKaRa) in a standard PCR reaction. PCR products were digested with the SURVEYOR nuclease cleaving single base pair mismatches in heteroduplex DNA (TILL et al. 2004). The undigested wild-type band was at 989 bp, whereas any sample containing the SNP displayed an undigested band at 989 bp and the digestion products at 719 bp and 217 bp. In addition an independent T-DNA line disrupting the identified gene was analyzed (SAIL_400_A01) and showed the same pollen tube overgrowth phenotype as the EMS allele (Figure 2).

Supporting references:

HUCK, N., J. M. MOORE, M. FEDERER, U. GROSSNIKLAUS, 2003 The Arabidopsis mutant feronia disrupts the female gametophytic control of pollen tube reception. Development **130:** 2149-2159

GIRAUT, L., M. FALQUE, J. DROUAUD, L. PEREIRA, O.C. MARTIN, C. MÉZARD, 2011 Genome-wide crossover distribution in Arabidopsis thaliana meiosis reveals sex-specific patterns along chromosomes. Plos Genet. **7:** e1002354

LISTER, C. AND C. DEAN, 1993 Recombinant inbred lines for mapping RFLP and phenotypic markers in Arabidopsis thaliana. Plant J. **4:** 745-750.

R DEVELOPMENTAL CORE TEAM, R: A language and environment for statistical computing. Vienna.

TILL, B. J., C. BURTNER, L. COMAI AND S. HENIKOFF, 2004 Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res. **32:** 2632-2641

Table S1 Simulation of a binomial distribution to determine the minimal sequencing coverage needed to differentiate a 1:1 from a 3:1 segregation of SNPs.

n	x	α F(1:1)	β 1-F(1:3)
20	5	0.0207	0.3828
30	10	0.0494	0.1057
40	14	0.0403	0.0544
50	18	0.0325	0.0287
60	23	0.0462	0.0075

Simulation of a binomial distribution to determine the minimal sequencing coverage needed to differentiate a 1:1 from a 3:1 segregation of SNPs, with 1st (α) and 2nd class (β) mistakes being <0.05. Π – values for H_0: Π = 0.5; H_A: Π= 0.25. The optimum is found at a sample size of 50. α = F~Binomial; β = 1-F~Binomial; n = sequence coverage; x = highest occurence of variant SNP to reject H0.

Table S2 Overview of the SNP data after mapping the reads to the *Arabidopsis thaliana* genome

Total SNPs	**2337**
homozygous	521
heterozygous	1816
non-synonymous	122
synonymous	69
mean coverage of heterozygous SNPs	**57.3**

SNPs were called by using the DiBayes algorithm with high stringency settings and analyzed with a custom software pipeline.

Table S3 SNP data of the resequenced EMS-mutagenized genome of *tun-1* (Col-0 accession).

All data presented has been obtained using the DiBayes algorithm for SNP calling with high stringency setting. The data is presented in a separate Excel file and available for download at http://www.genetics.org/content/suppl/2012/05/26/genetics.112.141341.DC1.

Table S4 Overview of the ten remaining 1:1 class candidate SNPs after χ^2 goodness of fit test for an expected pattern of co-segregation

chromosome	position	coverage	ratio	χ^2	cM (covered by the linkage group)
1	4891806	54	0.44	9.27	6.831
1	***5672441***	***68***	***0.49***	***7.29***	***4.673***
1	15198294	84	0.42	4.47	0.009
1	17143759	78	0.55	9.42	0.001
2	4905462	74	0.46	4.86	0.011
2	4905990	55	0.42	4.79	0.013
3	13191694	65	0.42	8.67	0.004
3	13191833	71	0.48	8.81	0.008
4	4003316	94	0.52	7.92	0.004
5	12012845	58	0.53	5.21	0.001

To test which SNPs differ from the expected pattern of co-segregation in a linkage group (five neighboring SNPs), a χ^2 goodness of fit test was applied (for $\alpha < 0.05$ and df = 4, $\chi^2 = 9.49$; en.wikipedia.org/wiki/Chi-squared_distribution). All SNPs with a $\chi^2 > 9.49$ were discarded (n=108), because they differ significantly from the genetic expectations, whereas all 10 SNPs with a $\chi^2 < 9.49$ have been kept and are shown in the above table. The line in bold and italic corresponds to the *tun-1* SNP.

EXTENDED EXPERIMENTAL PROCEDURES

Plant Material

Arabidopsis thaliana accessions Columbia-0 (Col), Landsberg *erecta* (L*er*), C24, WS or Nossen (No) were used as wild-type controls depending on the mutant investigated. The reporter lines used are listed in Table S3 with the corresponding reference. The *pAtRP-S5A:uidA* reporter was reconstructed in a different vector backbone (pCAMBIA1391Z) using the same promoter as originally described (Weijers et al., 2001) and transformed into the L*er* accession. The mutants used are listed and referenced in Table S3. All mutants were homozygous for a recessive null mutation except for *ddm1-2/DDM1*, *met1-3/MET1*, *msi1-2/MSI1* and *mea-2/MEA*. For the latter, heterozygous mutants were genotyped in a segregating population. For *ddm1-2* a cleaved amplified polymorphic sequence (CAPS) marker (gift from E. Richards) was used. A 100 bp amplicon produced with the primers 5′-gttggacagtgtggta aattccgct-3′ and 5′-gagctacgagccatgggtttgtgaaacgta-3′ (Tm 56°C, 40 cycles) was ethanol precipitated and digested with RsaI (NEB, Ipswich CA, USA) for 3 hr at 37°C. *msi1-2/MSI1*, *mea-2/MEA*, *met1-3/MET1* plants were genotyped as described (Baroux et al., 2006; Hennig et al., 2003; Saze et al., 2003). For the *htr4-1* mutant (line N582765, T-DNA insertion in the second exon of *At4g40030*) (Okada et al., 2005), abolished transcription was verified by RT-PCR using the primers 5′-tggctcgtaccaag-caaaccgctcg-3′ and 5′-acggactagcctctgaaatggcagtt-3′ (Tm 62°C, 35 cycles), targeting exon2 and exon 3, respectively. Equal loading was controled by amplifying *ACTIN11* mRNA as described (Baroux et al., 2006).

Preparation of Embryonic cDNA Libraries and Sequencing

For embryo isolation 3–5 siliques resulting from crosses between L*er* wild-type or *kyp-2/kyp-2* (L*er*) mutant mothers and Col wild-type fathers were harvested 2.5 days (2–4 cell embryos) and 4 days (globular) after pollination. Seeds were dissected and immersed in 20 μl isolation buffer (first-strand cDNA synthesis buffer [Invitrogen], 1.6 U/μl RNase Out [Invitrogen], 1 mM DTT), in a round-bottom 2 ml Eppendorf tube. Seeds were gently crushed with a plastic pestle to release the embryos. 400 μl isolation buffer was added to the extract and 5 x 50 μl droplets were placed on 6-well printed slides previously coated with 1% BSA. 50 μl fresh isolation buffer was placed on the remaining well for washing the isolated embryos (see below).

One slide was placed on an inverted microscope (Nikon TMS) and the droplets were screened at magnification 10x. Embryos were isolated with a siliconized, manually drawn, and freshly BSA-coated glass capillary fixed to a micromanipulator and linked to a 200 μl pipette with a rubber tube. The calibration wheel of the pipette was used to create a slight vacuum in the capillary to collect the embryos with as little solution as possible. The embryos were released in a clean drop of buffer and collected again (in ca. 2–3 μl) before release in 100 μl RNA extraction buffer. For profiling of wild-type transcriptomes (L*er* x Col) 28 and 4 embryos at the 2–4 cell stage and globular stage, respectively, were isolated. 25 mutant embryos from *kyp-2* x Col crosses were isolated. Plants were grown, crossed and harvested at the same time. RNA extraction was performed using the PicoPure RNA Isolation Kit (Arcturus) according to the manufacturer's instructions. Quality of the total embryonic RNA was assessed using Agilent RNA 6000 Pico Kit on the Agilent 2100 BioAnalyzer (Agilent Technologies, Germany) and we estimated a yield of 700 pg to 4 ng total RNA in each sample. 300–700 pg of total RNA was amplified using the WT-Ovation Pico RNA Amplification System (NuGEN Technologies, USA). The amplification technology is inspired from Philipps and Eberwine (1996) and performs a linear isothermal amplification of mRNA species. Unlike PCR-based exponential amplification, this linear amplification approach is carried out by replication of only the original transcripts, not replication of copies. Amplification of our samples produced 6–10 μg single-stranded cDNA. To create the second strand, poly(A) tails were added to 1 μg of the amplified cDNA library (10 pMol DNA ends) using 20U of Terminal Transferase (New England Biolabs, USA) and 0.2 mM dATPs in the provided buffer. The second strand was amplified during 1 PCR cycle on a thermal cycler (30″ at 95°C, 2 min at 50°C, 20 min at 72°C) using the Ex Taq Polymerase (TaKaRa, Japan) and oligo(dT)$_{12\text{-}18}$ Primer (Invitrogen).

Typically 200–500 ng of cDNA was used for SOLiD system's express fragment library preparation. Using the Covaris S2 system (Covaris, Inc.), cDNA (0.5–3 kb) was sheared into 80–130 bp short fragments according to the protocol. The ends of the target DNA were repaired and subsequently ligated to SOLiD P1 and P2 adaptors. After ligation the library was enriched by PCR (7 cycles) and a size selecting gel was run to remove any short fragments. The resulting ligated population was the SOLiD Fragment Library ready for emulsion PCR. Emulsion PCR reactions were performed according to the manufacturer's recommendation (Applied Biosystems, USA) by mixing 170 pg libraries with 0.8 billion 1 mm-diameter beads with P1 primers (ABI) covalently attached to their surfaces. Sequence reads of 50 bases length were generated by SOLiD v3 (Applied Biosystems, USA).

Allele-Specific Transcriptome Analysis

50 base reads generated by SOLiD v3 were aligned to the TAIR8.0 version of the *Arabidopsis* Col genome using the SOLiD System Analysis Pipeline Tool (Corona Lite 4.0r2.0, Applied Biosystems, USA), allowing up to 4 color-space mismatches with the additional rule "count valid adjacent errors as single errors." For transcriptome profiles, reads were excluded if they mapped to more than one genomic position, mapped at splice junctions or were partially overlapping. Consequently, full-length reads uniquely mapping inside a transcript were taken.

Arabidopsis L*er* SNPs (ftp://ftp.Arabidopsis.org/Polymorphisms/Ecker_ler.homozygous_snp.txt) (Borevitz et al., 2007) were used to identify reads matching L*er* sequences. We only used SNPs that were biallelic and have exactly one defined allele for Col and one

defined allele for L*er*. Under such constraints, we removed 24 SNPs from the published list for our calling procedure. To assign the reads to a L*er* or Col allele, we used the SNP calling pipeline (consensus caller) of Corona Lite 4.0r2.0. This procedure delivers the number of reads matching the reference sequence and the alternative sequence when at least 3 alternative alleles could be found. For each base change (alternative sequence) at a given position, probability and confidence scores are calculated and are used by the calling algorithm to categorize heterozygous and homozygous SNPs. In addition, the alternative sequence must be met by at least 3 reads with independent starting points. Consequently, in our case SNPs sequenced from one parent only (for instance L*er* if Col is used as the reference) would not be detected. To correct for this, we ran the SNP calling procedure twice, using a reciprocal set-up where the reference sequence was either L*er* or Col. To this aim, we generated two references with the new SNP list: one with the Col allele at all 304,978 genomic positions and one with the L*er* allele. Next, we checked for inconsistencies and only SNP calls showing the exact predicted L*er* or Col base were taken, SNP calls showing a different base were excluded (32 SNPs). To calculate the coverage at each SNP position, the results from the reciprocal SNP calling were merged, choosing the read count with highest number of unique start points then highest coverage. To calculate the allele-specific coverage (Col, L*er*), the coverage per SNPs per transcript was added following a procedure to eliminate redundant read counts, when those were covering 2 (or more) SNP. The procedure consisted in (1) sorting the SNPs by highest number of unique start points first, highest number of coverage second, (2) SNPs were interrogated, starting with the highest covered SNP, for their position and only SNPs more than 49 bp apart from a previously chosen SNP are kept, (3) we summarized Col- and L*er*-specific coverage and start points of all SNPs kept. When SNPs matched an annotation with two or more entries (transcript version), the entry with the highest coverage was used for calculating the transcript level.

Analysis of the Parental Distribution of Embryonically Expressed Genes Covered by SOLiD Reads

We assume that all gene expression patterns are either biparental, uniparental maternal, or uniparental paternal. A gene with both maternal and paternal transcripts is necessarily biparentally expressed. However, when transcripts from a single parent are detected but only a few reads have been sequenced, it is not immediately clear whether the gene is uni- or biparentally expressed, and the probability of missing one parental contribution in the sample needs to be considered. The relative probability of being uni- versus biparentally expressed was obtained using the following analysis. We note q the proportion of maternal transcripts: q = 0 for only paternal expression, q = 1 for only maternal expression and 0 < q < 1 for biparental expression. We build a model to adjust the distribution of q. We note θ_{pat}, θ_{mat}, θ_{bip} the proportion of genes that are paternally, maternally, or biparentally expressed, respectively. We assume that q values for genes that are biparentally expressed is Beta distributed (with parameters θ_a and θ_b). We note m_i and p_i the observed number of paternal and maternal transcripts for gene i. We note **m** and **p** the vector of all m_i and p_i and θ the vector of parameters to be estimated. The likelihood of the data can then be written:

$$L(\mathbf{m},\mathbf{p}\mid\boldsymbol{\theta})=\prod_i\begin{cases}\theta_{pat}+\theta_{bip}\int_0^1\beta(\theta_a,\theta_b;x)\mathrm{B}(p_i,x;0)dx & \text{if } m_i=0\\ \theta_{bip}\int_0^1\beta(\theta_a,\theta_b;x)\mathrm{B}(p_i+m_i,x;m_i)dx & \text{if } m_ip_i>0\\ \theta_{mat}+\theta_{bip}\int_0^1\beta(\theta_a,\theta_b;x)\mathrm{B}(m_i,x;m_i)dx & \text{if } p_i=0\end{cases}\quad(1)$$

where $\beta(\theta_a, \theta_b ; x)$ denotes the probability to draw x in a Beta distribution with parameters θ_a and θ_b and where B(*n, x; k*) denotes the probability to draw k success among n trials with a probability of success x (i.e., in a binomial distribution with parameters n and x). For consistency, we assume that the mode of q distribution for biparentally expressed genes is not 0 or 1, which entails that $\theta_a > 1$ and $\theta_b > 1$. If no maternal transcripts are present, the gene may be paternally (with probability θ_{pat}) or biparentally expressed (with probability θ_{bip}). In the latter case, the probability to not sample any maternal transcript is evaluated by the corresponding binomial distribution integrated over the distribution of q within the biparentally expressed genes (the Beta distribution). This leads to the first line of Equation 1. Similar reasoning yields the two other lines in the equation. The relative probability of being uni- versus biparentally expressed can then be calculated. A gene with no maternal transcript is exclusively paternally expressed with probability

$$P_i=\frac{1}{1+\int_0^1\beta(\hat{\theta}_a,\hat{\theta}_b;x)\mathrm{B}(p_i,x;0)dx}\quad(2)$$

and biparentally expressed with probability $B_i = 1 - P_i$. Similarly, a gene with no paternal transcript is exclusively maternally expressed with probability

$$M_i=\frac{1}{1+\int_0^1\beta(\hat{\theta}_a,\hat{\theta}_b;x)\mathrm{B}(m_i,x;m_i)dx}\quad(3)$$

and biparentally expressed with probability $B_i = 1 - M_i$. In Equations 2 and 3, the hat denotes the maximum likelihood estimates of the parameters that have been obtained by maximizing $L(\mathbf{m},\mathbf{p}\mid\boldsymbol{\theta})$. We performed this analysis independently for the three samples

analyzed (2–4 cell wild-type embryos, globular embryos, and 2–4 cell *kyp/KYP* embryos). P_i, M_i, and B_i values for all genes are provided in Table S2.

To obtain a global view of the variations of the parental expression between the samples analyzed, we computed transition matrices. More precisely, we computed the overall proportion of genes being maternally, biparentally or paternally expressed in sample 1 becoming maternally, biparentally or paternally expressed in sample 2. For instance the proportion of maternally expressed genes in 2–4 cell wild-type embryos (WT24) becoming biparentally expressed in Globular embryos (WT Glob) was computed as:

$$\sum_{i=1}^{\#\ genes} M_{i(\mathrm{WT24})} B_{i(\mathrm{WTGlob})}, \tag{4}$$

where M_i and B_i are given by Equations 2 and 3. Thus, these transition matrices account for the uncertainty of assigning each gene to maternal, biparental, or paternal categories. Two matrices were computed, for (2–4 cell wild-type -> globular) and (2–4 cell wild-type -> 2–4 cell *kyp/KYP*) transitions.

Histochemical Detection of the uidA Reporter Gene Product (GUS Staining)

Developing siliques were cut longitudinally and fixed in ice-cold 90% acetone for < 1 hr at −20°C. After washing three times with 100 mM phosphate buffer (100 mM Na2HPO4, 100 mM NaH_2PO_4), the tissue was immersed in staining solution (0.1% Triton X-100, 10 mM EDTA, 0.5 mM Ferrocyanide, 0.5 mM Ferricyanide and 4 mM 5 -bromo-4-chloro-3-indolyl-beta-d-glucuronic acid cyclohexyl-ammonium salt (X-gluc, Biosynth AG, Staad, CH) in 100 mM phosphate buffer) and vacuum-infiltrated for 5 min. The staining reaction was carried out for 2 days at 37°C except for ET1041 (4 days) and *pGRP23:uidA* (2 hr). The staining solution was removed and the samples mounted in clearing solution (40 g of chloralhydrate (Sigma, Steinheim, DE) dissolved in 5 ml Glycerol, 1 ml Lactic acid, and 10 ml water).

Quantification of Paternal Reporter Gene Expression following Crosses

Wild-type and mutant plants were pollinated with the marker lines listed in Table S3, 2 days after emasculation. In the case of *msi1-2/MSI1* and C24, however, pollination was carried out 1 day after emasculation to limit the formation of autonomous seed development (Hennig et al., 2003). Developing siliques were harvested at different days after pollination and stained as described above. The seeds showing GUS staining were scored under a Leica HC microscope (Leica Microsystems, Wetzler GMBH, DE) or Zeiss Axio Imager microscope (Carl Zeiss MicroImaging GmbH, DE). Imaging was done with a CCD camera (Magnafire - Optronics, Goletta, USA) and images edited using Graphic Converter (lemkeSOFT). The seeds were scored according to four classes corresponding to the following developmental stages: zygote to 4 cell stage (embryo proper), octant to 16 cell stage, globular stage, and heart stage. In the case of *msi1-2/MSI1* and C24, the earliest class (zygote to 4 cell stage) was not considered because of potential bias due to parthenogenetic embryo development in the *msi1* mutant (Guitton and Berger, 2005). Average and standard error of the relative proportion of GUS staining seeds in the different developmental classes were calculated from independent biological replicates (Table S4). For each replicate, wild-type and mutant samples were processed in parallel (cross, GUS staining, scoring) to ensure comparable results. Differences in the number of GUS-positive seeds, either between wild-type and mutant samples or between consecutive developmental stages, were tested for statistical significance using the two-tail Fisher's exact test (http://www.langsrud.com/fisher.htm). The details of the scoring results are given in Table S4.

Allele-Specific RT-PCR

Total RNA was isolated (Trizol reagent, Invitrogen) from siliques collected at 2, 3, 4, and 5 days after pollination and treated with DNase I (Invitrogen). cDNA was produced using SuperSript III (Invitrogen) and polyT primers following the manufacturer's instructions. *AGP18* (*At4g37450*) CDS shows a single-nucleotide polymorphism (SNP) between the L*er* and Col ecotypes. An LNA-modified primer was designed (AGP18-Col-R: 5′-gcagttggagttttcgccggagc+c-3′) ("+" before the base indicates the LNA modified base) that, together with a nonspecific primer (AGP18-F5′-ggccaatctcctatctcttctccga-3′) specifically amplifies the Col allele. In a cross between a wild-type L*er* or *kyp-2* female and a Col pollen donor, this primer pair specifically detects the paternal Col allele following 40 cycles with 62°C annealing temperature. As an internal control, both parental alleles were amplified using AGP18-F and AGP18-R (5′-gcagttggacttttgccggagct-3′). Similarly, a SNP in *GRP23* (*At1g10270*) allowed distinguishing the C24 maternal and L*er* paternal alleles in crosses between C24 wild-type or *msi1-2* females and L*er* pollen donors. GRP23-Ler-R (5′-cggtggctgttgccctgccgt+c-3′) and GRP23-F (5′- gcaggtcaaacagcaggaggag-3′) specifically amplified the paternal L*er* allele using 39 cycles with 63°C annealing temperature. Amplification of both parental alleles was obtained using GRP23-F and GRP23-R (5′-cggtggctgttgccctgccgtt-3′) primers. Control PCR reactions were carried out on samples without reverse transcriptase to confirm the absence of genomic DNA. RT-PCR analysis of the *PHE1* gene was done using primers described previously (Köhler et al., 2005) and 36 cycles of amplification with 62°C annealing temperature.

Profiling of Ovule Small RNAs

Mature ovules were collected using a custom micro-pump (Peiffer et al., 2008), avoiding contamination by placenta and silique valve tissues. Total RNA was extracted using the Trizol reagent according to the manufacturer's instructions. A small RNA library was prepared and sequenced using current Illumina Genome Analyzer protocols (http://www.illumina.com). Approx. 10 millions reads were obtained, which were filtered for adaptator removal, size (reads between 19 and 27 nt were conserved), mismatches, and redondancy using custom PERL scripts and the MEGA-BLAST algorythm (http://blast.ncbi.nlm.nih.gov/). The library finally contained 1,193,800 unique reads, which were mapped against *Arabidopsis* Col reference sequence (TAIR 8) using the BOWTIE software (http://bowtie-bio.sourceforge.net/index.shtml) (Langmead et al., 2009). The same prodecure was used to analyze a 995,650 read inflorescence small RNA library, sequenced by B. Meyers' group (library code: FLR, http://mpss.udel.edu/at/tiny_library.php?lib=1) (Lu et al., 2005). Mapping and occurrence informations, after normalization of both libraries to 1 million, were compiled and graphical displays were produced using R. For relative CDS and repeat targeting analysis, two different reference databases (indexes), CDS and Repeats, were built based on TAIR 8 information (ftp://ftp.Arabidopsis.org/home/tair/Genes/TAIR8_genome_release/) and repeat annotation available (ftp://ftpmips.helmholtz-muenchen.de/plants/cress/), and compared after mapping with BOWTIE.

Whole-Mount Immunolocalization

All antibodies were obtained from Abcam (Cambridge, UK). The H5 (#ab24758) raised against the active form of PolII specifically targets the CTD of the main sub-unit of PolII when it is phosphorylated on serine 2, which occurs during transcript elongation. As shown previously (Palancade and Bensaude, 2003), the CTD phosphorylation pattern is modified as PolII engages in transcript elongation and the H5 antibody detects Pol II molecules during active transcription. Chromatin analysis was performed using H3K9me2 (#ab1220) or H3K9me3 (#ab71999) antibodies. Young siliques were fixed for 3 to 4 hr in 4% paraformaldehyde in 1xPBS with 2% Triton, washed twice in 1xPBS. Young seeds were dissected and embedded in acrylamide on slides as described (Bass et al., 1997). Samples were digested with 1% driselase, 0,5% cellulase, 1% pectolyase (all from Sigma) in 1XPBS with 1%BSA for 25 min to 1 hr at 37°C, subsequently rinsed 3 times in 1xPBS, and permeabilized for 1 to 2 hr in 1xPBS with 2% Triton. The primary antibodies were applied at the following dilutions: 1:400 for H3K9me2 and H3K9me3; 1:200 for H5, overnight at 4°C. The slides were washed day-long in 1xPBS, 0,2% Triton, and coated with secondary antibody (Alexa Fluor 488 conjugate, Molecular Probes) used at a 1:400 dilution. After washing in 1xPBS; 0,2% Triton for a minimum of 6 hr, the slides were incubated with DAPI (1ug/ml in 1xPBS) for 1h, washed for 1h in 1xPBS, and mounted in PROLONG medium (Molecular Probes). Images were captured on a laser scanning confocal microscope (Leica SP2) equipped for DAPI (405nm) and FITC (488 nm) excitation and either 40X or 63X objectives. Maximum-intensity projections of selected optical sections were generated, and edited using Graphic Converter (lemkeSOFT).

SUPPLEMENTAL REFERENCES

Acosta-Garcia, G., and Vielle-Calzada, J.P. (2004). A classical arabinogalactan protein is essential for the initiation of female gametogenesis in *Arabidopsis.* Plant Cell *16*, 2614–2628.

Bass, H.W., Marshall, W.F., Sedat, J.W., Agard, D.A., and Cande, W.Z. (1997). Telomeres cluster de novo before the initiation of synapsis: a three-dimensional spatial analysis of telomere positions before and during meiotic prophase. J. Cell Biol. *137*, 5–18.

Blevins, T., Rajeswaran, R., Shivaprasad, P.V., Beknazariants, D., Si-Ammour, A., Park, H.S., Vazquez, F., Robertson, D., Meins, F., Jr., Hohn, T., et al. (2006). Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res. *34*, 6233–6246.

Cao, X., and Jacobsen, S.E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc. Natl. Acad. Sci. USA *99* (*Suppl 4*), 16491–16498.

Chan, S.W., Henderson, I.R., Zhang, X., Shah, G., Chien, J.S., and Jacobsen, S.E. (2006). RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in *Arabidopsis.* PLoS Genet. *2*, e83.

Colon-Carmona, A., You, R., Haimovitch-Gal, T., and Doerner, P. (1999). Technical advance: spatio-temporal analysis of mitotic activity with a labile cyclin-GUS fusion protein. Plant J. *20*, 503–508.

Ding, Y.H., Liu, N.Y., Tang, Z.S., Liu, J., and Yang, W.C. (2006). *Arabidopsis* GLUTAMINE-RICH PROTEIN23 is essential for early embryogenesis and encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III. Plant Cell *18*, 815–830.

Phillips, J., and Eberwine, J.H. (1996). Antisense RNA Amplification: A linear amplification method for analyzing the mRNA population from single living cells. Methods *10*, 283–288.

Ferreira, P.C., Hemerly, A.S., Engler, J.D., van Montagu, M., Engler, G., and Inze, D. (1994). Developmental expression of the *Arabidopsis* cyclin gene cyc1At. Plant Cell *6*, 1763–1774.

Gross-Hardt, R., Kägi, C., Baumann, N., Moore, J.M., Baskar, R., Gagliano, W.B., Jurgens, G., and Grossniklaus, U. (2007). LACHESIS restricts gametic cell fate in the female gametophyte of *Arabidopsis.* PLoS Biol. *5*, e47.

Grossniklaus, U., Vielle-Calzada, J.P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal control of embryogenesis by MEDEA, a Polycomb group gene in *Arabidopsis.* Science *280*, 446–450.

Guitton, A.E., and Berger, F. (2005). Loss of function of MULTICOPY SUPPRESSOR OF IRA1 produces nonviable parthenogenetic embryos in *Arabidopsis.* Curr. Biol. *15*, 750–754.

Jackson, J.P., Johnson, L., Jasencakova, Z., Zhang, X., PerezBurgos, L., Singh, P.B., Cheng, X., Schubert, I., Jenuwein, T., and Jacobsen, S.E. (2004). Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana.* Chromosoma *112*, 308–315.

Jeddeloh, J.A., Stokes, T.L., and Richards, E.J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat. Genet. *22*, 94–97.

Köhler, C., Page, D.R., Gagliardini, V., and Grossniklaus, U. (2005). The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nat. Genet. *37*, 28–30.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., et al. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. EMBO J. *23*, 4286–4296.

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis.* Genes Dev. *19*, 2030–2040.

Saze, H., Mittelsten Scheid, O., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat. Genet. *34*, 65–69.

Sung, S., He, Y., Eshoo, T.W., Tamada, Y., Johnson, L., Nakahigashi, K., Goto, K., Jacobsen, S.E., and Amasino, R.M. (2006). Epigenetic maintenance of the vernalized state in *Arabidopsis thaliana* requires LIKE HETEROCHROMATIN PROTEIN1. Nat. Genet. *38*, 706–710.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007). *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genet. *3*, e86.

Weijers, D., Franke-van Dijk, M., Vencken, R.J., Quint, A., Hooykaas, P., and Offringa, R. (2001). An *Arabidopsis* Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene. Development *128*, 4289–4299.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol. *2*, E104.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716–719.

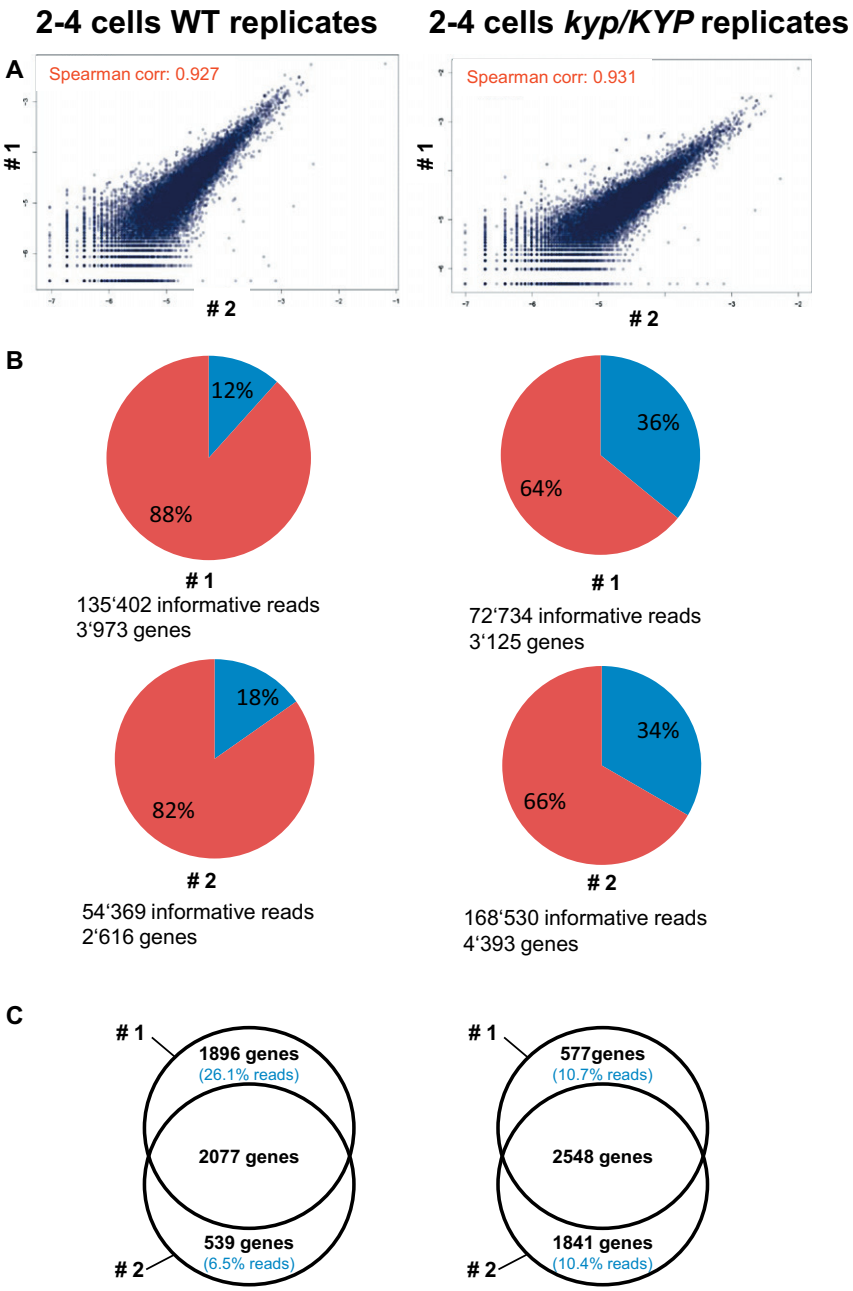


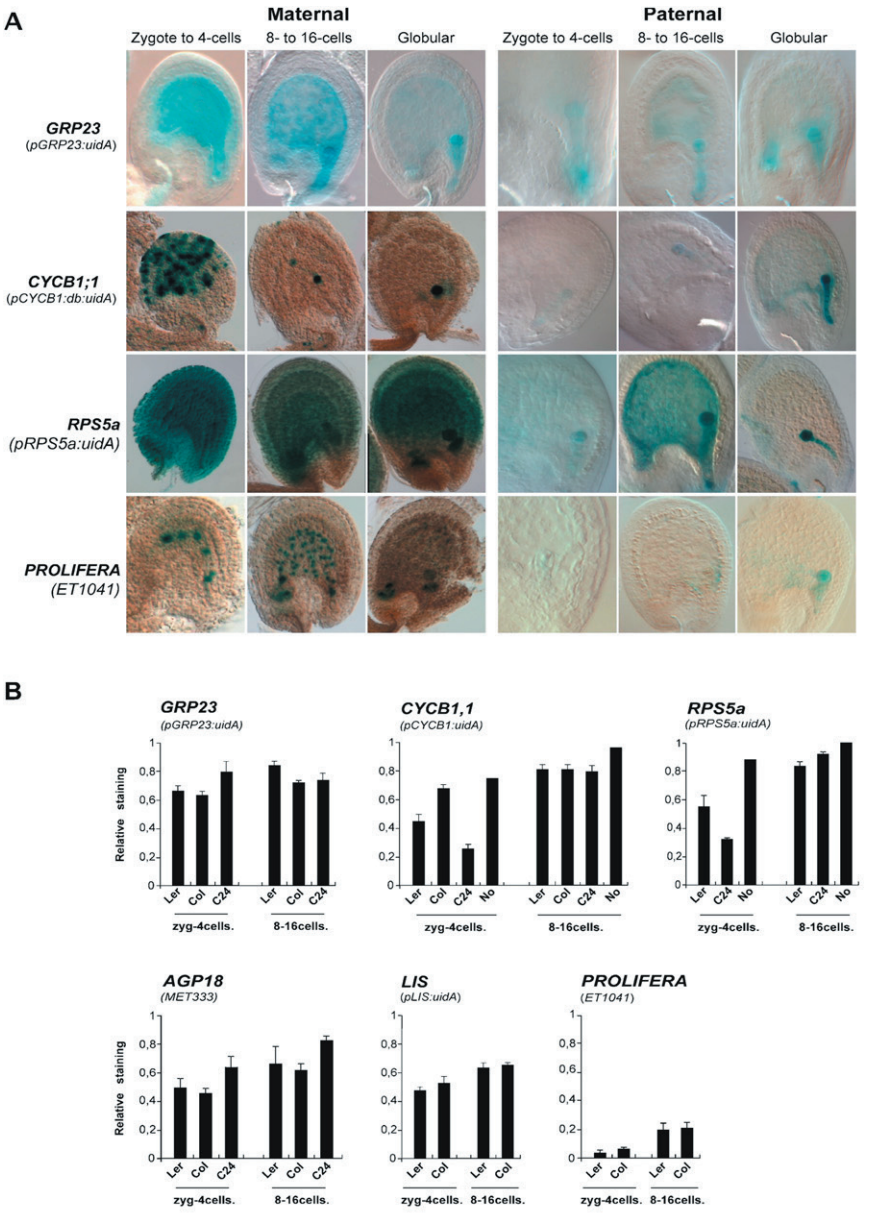
Figure S1. Replicate Transcriptome Profiling from Wild-Type and Mutant Embryos at the 2–4 Cell Stage, Related to Figure 1 and Figure 4

Allele-specific profiling of wild-type and *kyp/KYP* mutant embryos at the 2-4 cell stage was performed in two independent replicates (independent embryo collection, RNA extraction, amplification and sequencing).

(A) The replicate transcriptomes show overall good correlation. The transcript levels describing the embryonic transcriptomes of replicates #1 and #2 were plotted in log2 scale and the Spearman correlation is indicated.

(B) The parental read distribution is similar in each replicate. The distribution of paternal (blue) and maternal (red) reads is presented as in Figure 1. the number of informative reads and genes identified is indicated. Replicates #1 are those reproduced in Figure 1. Because of better coverage in the wild-type, the allele-specific analysis was carried out in depth for replicates #1.

(C) The genes identified by informative reads overlap well between replicates. Common genes are covered by the majority of reads (89.6%–93.5% of replicates #2). The Venn diagrams show the overlap between the informative transcriptomes (i.e., identified by informative reads covering SNP regions) of each replicates. Note that the lack of overlap does not necessarily indicate lack of expression but coverage below our detection threshold on the selected SNPs (i.e., genes appearing specific to one replicate can show read coverage in regions outside the SNPs). The genes common between replicates are covered by a majority of the reads: the 2077 common genes in WT replicates are covered by 93.5% reads of replicate #2 ; the 2548 common genes in *kyp/KYP* replicates are covered by 89.6% reads of replicate #2.



Cross	Zygote - 4 cells			8 - 16 cells			Globular		
	Relative staining	SE (rep)	n	Relative staining	SE (rep)	n	Relative staining	SE (rep)	n
ET1041 x Ler	**0,93**	0,00 (3)	289	**0,95**	0,02 (2)	53	**0,96**	0,02 (2)	35
Ler x ET1041	**0,03**	0,01 (6)	207	**0,19**	0,03 (7)	129	**0,40**	0,03 (7)	305
MET333 x Ler	**0,97**	0,01 (3)	64	**0,96**	0,01 (4)	56	**1,00**	0,00 (4)	32
Ler x MET333	**0,50**	0,06 (5)	214	**0,66**	0,13 (4)	128	**0,88**	0,03 (4)	110
pGRP23:uidA x Col	**0,96**	0,01 (2)	243	**0,67**	na (1)	27			
Col x pGRP23:uidA	**0,51**	0,00 (9)	293	**0,66**	0,00 (9)	368			

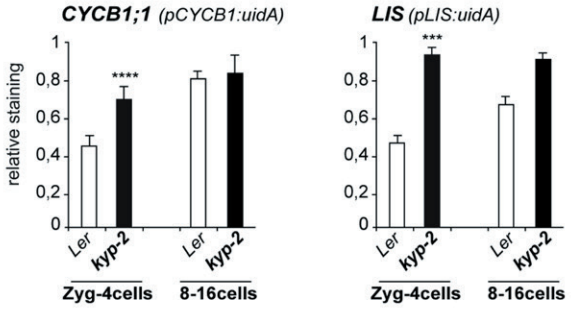
Figure S2. Maternal and Paternal Expression of Embryo Markers in Wild-Type, Related to Figure 3

(A) Differential expression of the indicated reporter transgenes (Table S3), when transmitted maternally (right panel) or paternally (left panel) was detected using histochemical detection of the *uidA* gene product (GUS) at different stages.

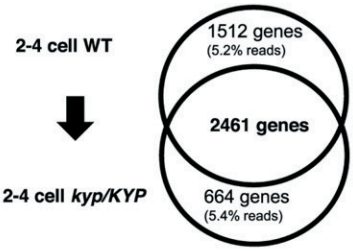
(B) Quantification of paternal expression in *Arabidopsis* accessions. The proportion of seeds showing GUS staining (relative staining) was compared in the progeny of crosses between wild-type plants from the indicated accessions: Landsberg *erecta* (L*er*), Columbia (Col), C24, Nossen (No), and the indicated marker lines. Error bars represent standard error (SE). Detailed analysis is provided in Table S4.

(C) Quantification of maternal and paternal expression. The proportion of seeds showing GUS staining (relative staining) was compared in the progeny of reciprocal crosses between wild-type and the indicated marker lines. SE, standard error. (rep), replicates. n, total number of seeds scored. See also Table S4.

A Paternal marker expression in wild-type and kyp mutant mother



B Transcriptome comparisons



C Class transitions among common genes

Number of genes

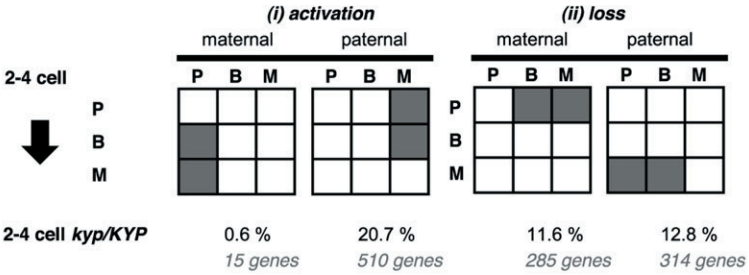
2-4 cell kyp/KYP \ 2-4 cell WT	P	B	M
P	5	174	111
B	8	1116	399
M	7	307	333

% of genes (heatmap)

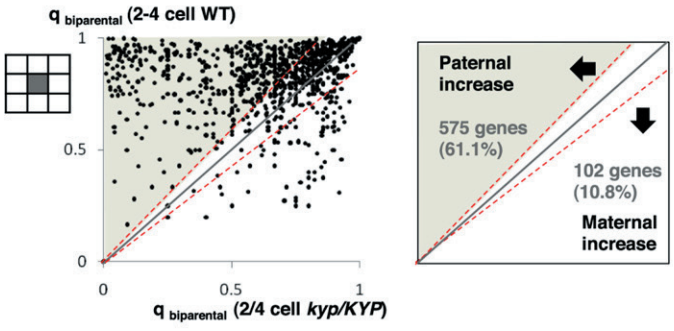
	P	B	M
P	0.2	7.1	4.5
B	0.3	45.4	16.2
M	0.3	12.5	13.6

P Paternal Class (q=0)
M Maternal Class (q=1)
B Biparental Class (0<q<1)

D Activation *vs* loss of one parental contribution



E Relative increase of paternal or maternal contribution among biparental-class genes (B > B transition)



F Summary: paternal alleles are preferentially upregulated in *kyp* embryos

	De novo detected	Relative increase (>10%)	Total
Paternal	510	575	1085 / 2461
Maternal	15	102	117 / 2461

Figure S3. Derepression of Paternal Markers and Changes of Parental Contributions Induced by a Maternal *kyp* Mutation, Related to Figure 4

(A) Embryos showing GUS staining were scored in the progeny of crosses between wild-type (L*er*) or *kyp-2* mutant females and the indicated paternal markers, at the developmental stages indicated above the graphs. Two-tailed Fisher's exact tests were carried out to assess the differences between wild-type and mutant samples. The levels of significance are indicated (*: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$; ****: $p < 0,0001$). Error bars represent SE. Detailed analysis is provided in Table S4.

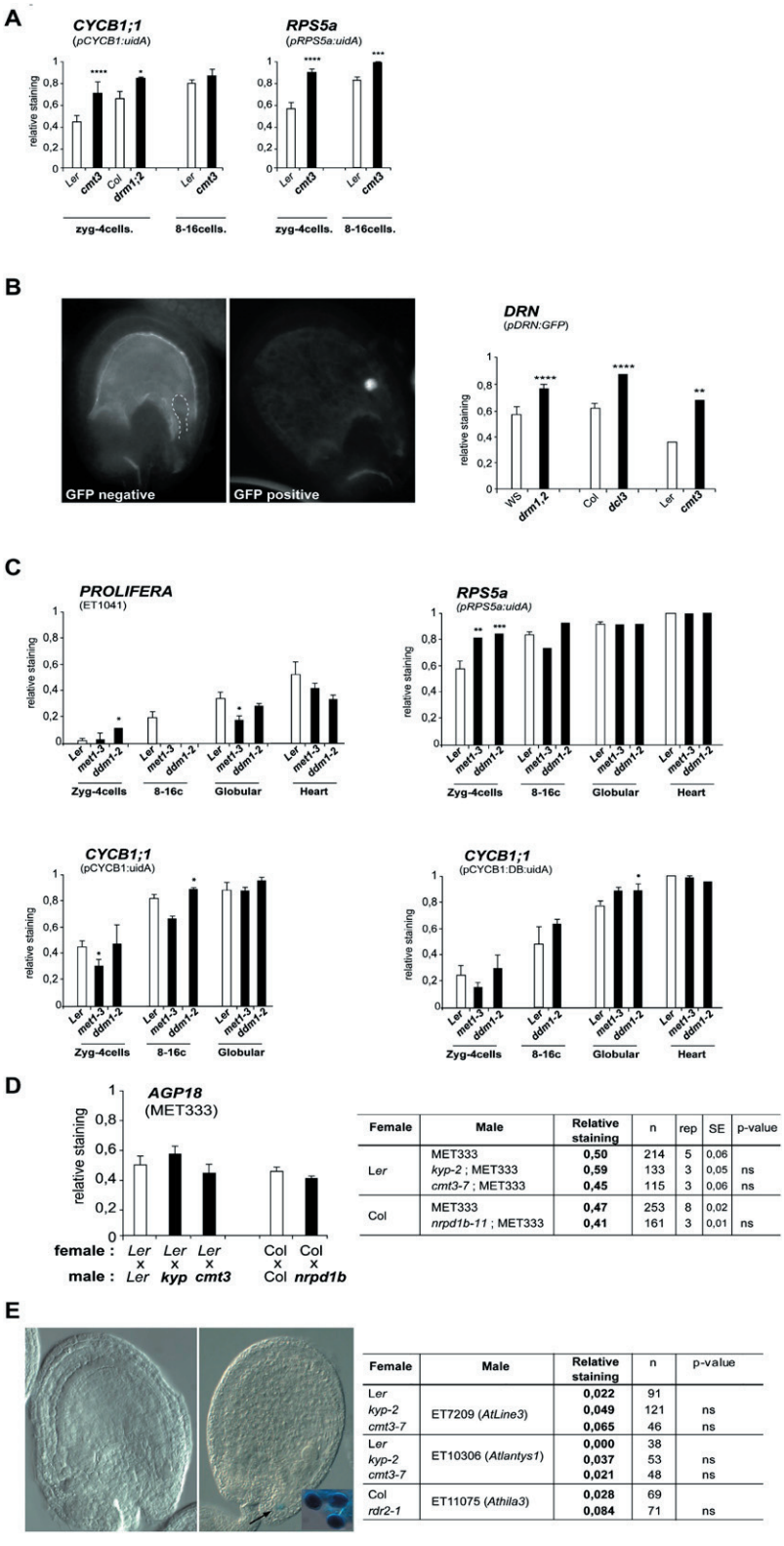
(B) Venn diagrams show the number of genes shared between the SNP-tagged transcriptomes (see Experimental Procedures) of 2-4 cell WT and 2-4 cell *kyp*/*KYP* embryos. Genes specifically detected in one sample only are covered by 5% reads, as indicated. Genes commonly detected in both samples, by contrast, are covered by the majority of reads (95%).

(C) The transition tables describe the changes for the common genes in the parental-class distributions (P, B, M, see legend in gray box) that were induced by a maternal *kyp* mutation (2-4 cell WT > 2-4 cell *kyp*/*KYP* transition). The transition matrices were calculated using parental-class probabilities (Pi, Bi, Mi) as described in Experimental Procedures.

(D) Several subsets of class transitions clearly illustrate cases of (i) activation/de novo expression and (ii) decreased expression /loss of one parental allele in our transcriptomes. Note that loss may correspond to a decrease in SNP coverage (see Experimental Procedures) falling below our detection threshold, rather than the absolute loss of transcript. The % indicated is the sum of the % genes in (C) falling in the gray transitions.

(E) A large fraction of common genes (45.4%) remain in the biparental class in both genotypes. The proportion of maternal transcripts was calculated for the 940 common genes effectively sequenced on both alleles. The calculated q values were plotted as indicated (left panel). The scheme (right panel) show the relative changes in parental contributions toward either higher paternal or higher maternal representation. Only genes with changes deviating from 10% (above/below dashed lines) were counted. 263 genes showed no or less than 10% change.

(F) The table summarizes the number of genes affected on either the maternal or the paternal allele in embryos inheriting a maternal *kyp* mutation. Shown are the genes with a novel contribution of one parental allele (de novo) in mutant embryos as reported in (D) and biparental-class genes showing a relative increase of one parental contribution as shown in E. Note that for the latter, only genes are counted that show a change >10% compared to in wild-type embryos (genes above/below the red lines).



Female	Male	Relative staining	n	rep	SE	p-value
Ler	MET333	**0,50**	214	5	0,06	
	kyp-2 ; MET333	**0,59**	133	3	0,05	ns
	cmt3-7 ; MET333	**0,45**	115	3	0,06	ns
Col	MET333	**0,47**	253	8	0,02	
	nrpd1b-11 ; MET333	**0,41**	161	3	0,01	ns

Female	Male	Relative staining	n	p-value
Ler	ET7209 (*AtLine3*)	**0,022**	91	
kyp-2		**0,049**	121	ns
cmt3-7		**0,065**	46	ns
Ler	ET10306 (*Atlantys1*)	**0,000**	38	
kyp-2		**0,037**	53	ns
cmt3-7		**0,021**	48	ns
Col	ET11075 (*Athila3*)	**0,028**	69	
rdr2-1		**0,084**	71	ns

Figure S4. Maternal *CMT3* and *DRM2*, but not *MET1* and *DDM1*, Control Early Expression of Paternally Transmitted Markers, Related to Figure 5

Additional control experiments showing maternal but not paternal effect of RdDM mutations and absence of effects on transposon enhancer trap markers.

(A) Embryos showing GUS staining were scored in the progeny of crosses between wild-type (Col or L*er*) or mutant females and the indicated paternal markers, at the developmental stages indicated above the graphs. Two-tailed Fisher's exact tests were carried out to assess the differences between wild-type and mutant samples. The levels of significance are indicated (*: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$; ****: $p < 0,0001$). Error bars represent SE. Detailed analysis is provided in Table S4.

(B) The *drm1-1,drm2-1* double mutant line contains a *GUS* transgene (see allele reference Table S3), a GFP embryo marker was therefore used to test the role of these genes on paternal gene activation. Paternal expression of *pDRN:GFP* observed 1.5 days after pollination is shown, longer exposure was used to visualize the young embryo devoided of GFP signal (outlined). The proportion of GFP positive embryos was scored in the progeny of the indicated crosses, 1.5 days after pollination. Stage distributions were similar among mutant and control populations, ranging from the zygote to octant stages. *dcl3-1*, Col, *cmt3-7* and L*er* females were also tested in parallel as controls. Two-tailed Fisher's exact tests were carried out to assess the differences between wild-type and mutant samples. The levels of significance are indicated (**: $p < 0,01$; ***: $p < 0,001$;****: $p < 0,0001$). Error bars represent SE. Detailed analysis is provided in Table S4. Marker and mutant lines are described in Table S3.

(C) Embryos showing positive GUS staining were scored in the progeny of crosses between wild-type (L*er*), *met1-3*/*MET1* or *ddm1-2*/*DDM1* females and the indicated paternal markers, at different developmental stages as noted below graphs. Two-tailed Fisher's exact tests were carried out to assess the differences between wild-type and mutant samples. The levels of significance are indicated (*: $p < 0,05$; **: $p < 0,01$). Error bars represent SE. Detailed analysis is provided in Table S4.

(D) Absence of paternal effects of *kyp* and *nrpd1b* mutations on paternally transmitted markers. The proportion of seeds showing GUS staining (relative staining) was compared in the progeny of crosses between wild-type females and male with or without the mutations, together with the paternal markers lines, as indicated. n, total number of seeds scored. Two-tailed Fisher's exact tests showed no significant differences (ns). Error bars represent SE.

(E) Paternally transmitted transposon enhancer trap lines are not reactivated in seeds derived from *kyp*, *cmt3* or *rdr2* mothers. Paternally transmitted enhancer trap line *ET7209* monitoring expression of *AtLine3* family transposon (Slotkin et al., 2009), as for *ET10306* (*Atlantys1*) and *ET11075* (*Athila3*), did not show expression in the early seed in most cases, as described (Slotkin et al., 2009) (left picture). Faint staining at the suspensor base was detected occasionally (right picture, arrow), and scored as GUS positive. Strong pollen staining was detected in all lines (inset in left picture). The table displays the proportion of early seeds (zygote to globular stages) showing GUS staining in crosses between wild-type or mutant females and three transposon GUS marker lines, as indicated. Staining in the embryo proper was never observed, and the percentage of seed with signal at suspensor base (see right picture) remained below 10% for all mutants tested. n, total number of seeds scored. Two-tailed Fisher's exact tests showed no significant differences (ns). Error bars represent SE.

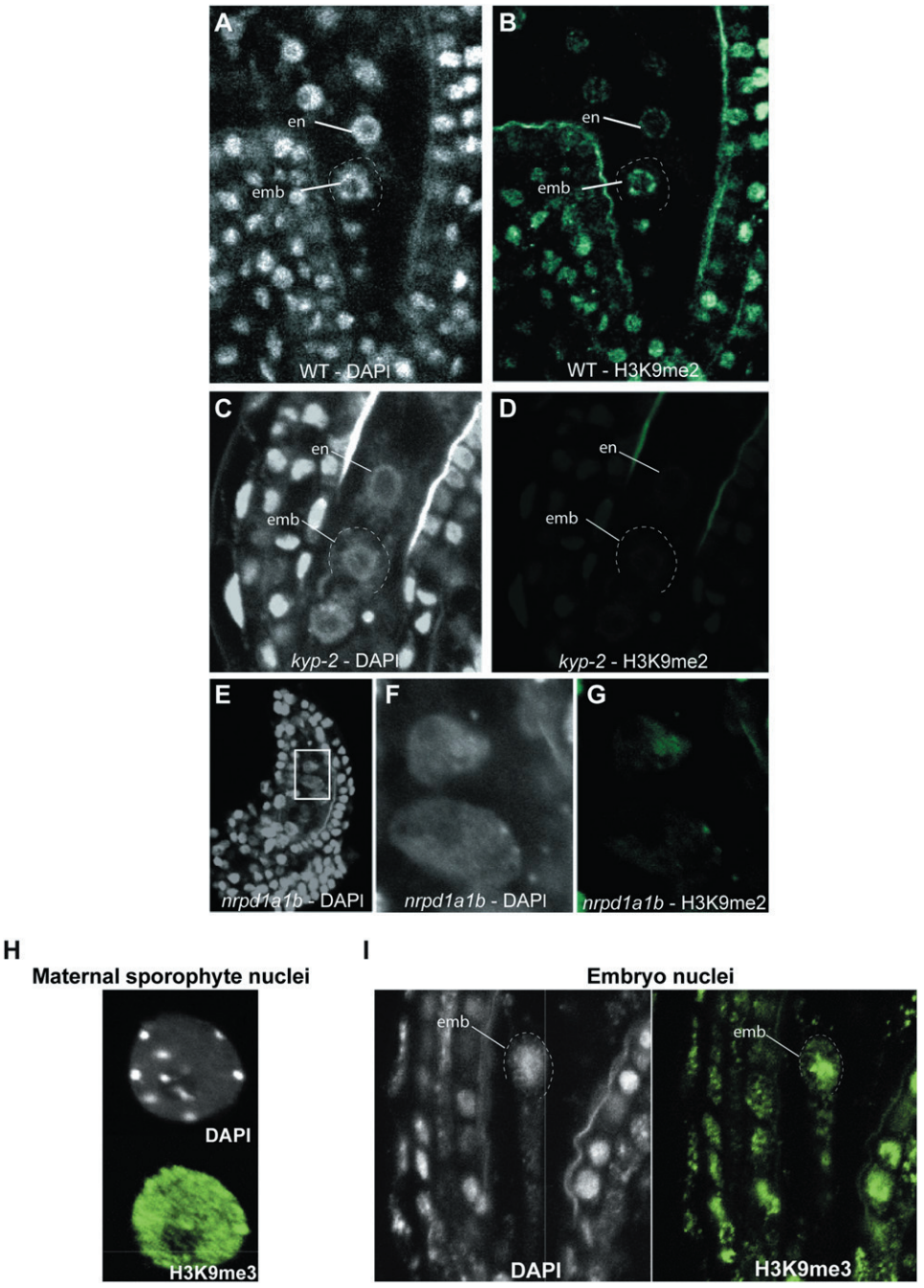
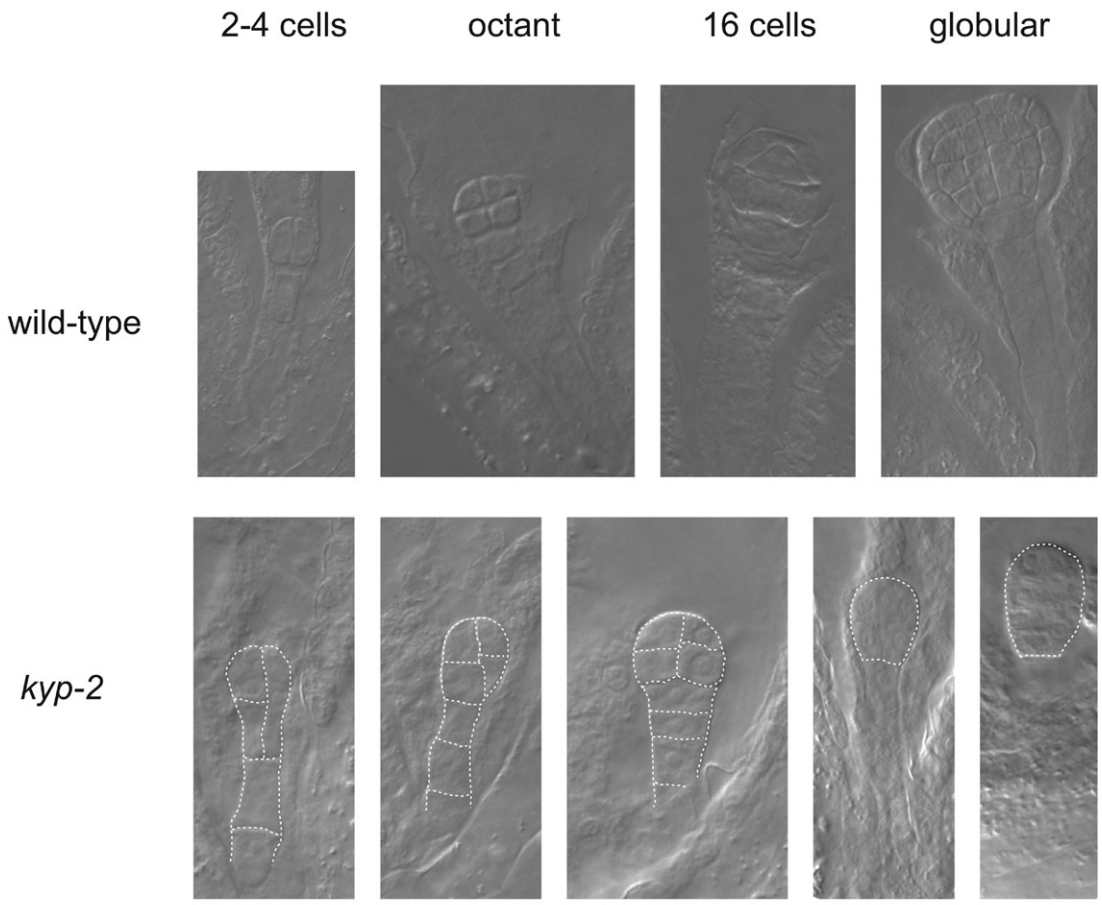


Figure S5. Control Immunodetection of H3K9me2 Showing Absence of Signals in *kyp* Embryos and Altered Distribution of *nrpd1* Mutant Embryos, by Contrast to H3K9me3 which Remain Unaffected in Both Mutants, Related to Figure 5

(A–G) Immunolocalization of histone H3 di-methylation on lysine 9 (H3K9me2) shows a strong signal in heterochromatic chromocenters in wild-type (A and B), while almost no signal in *kyp-2*, as described previously (Jackson et al., 2004) (C and D), and an altered, dispersed signal in *nrpd1a-1/nrpd1b-11* (*nrpd1a1b*) double mutant nuclei (E–G).

(H and I) Immunolocalization of histone H3 tri-methylation on lysine 9 (H3K9me3) showed that, by contrast to H3K9me2, H3K9me3 is present in both somatic (tegument, H) and embryonic chromatin (I). Specifically, it is homogenously distributed in euchromatin as described previously (Turck et al., 2007; Sung et al., 2006) in both wild-type, *kyp-2* and *nrpd1a-1/nrpd1b-11* backgrounds. en, endosperm. emb, embryo. The DNA was counterstained with DAPI.



	abnormal embryos	n	rep.
L*er*	2,6 % (+/- 0,2)	421	3
kyp-2	11,8 % (+/- 3,5)	260	3

Figure S6. Transient Patterning Defects in *kyp* Mutant Embryos, Related to Figure 4

Wild-type (L*er*) and *kyp-2* mutant early embryos were cleared using Herr's solution. Representative examples of the abnormal phenotypes observed in *kyp-2* are shown. Wild-type and abnormal phenotypical classes were scored in independent plants (rep.).

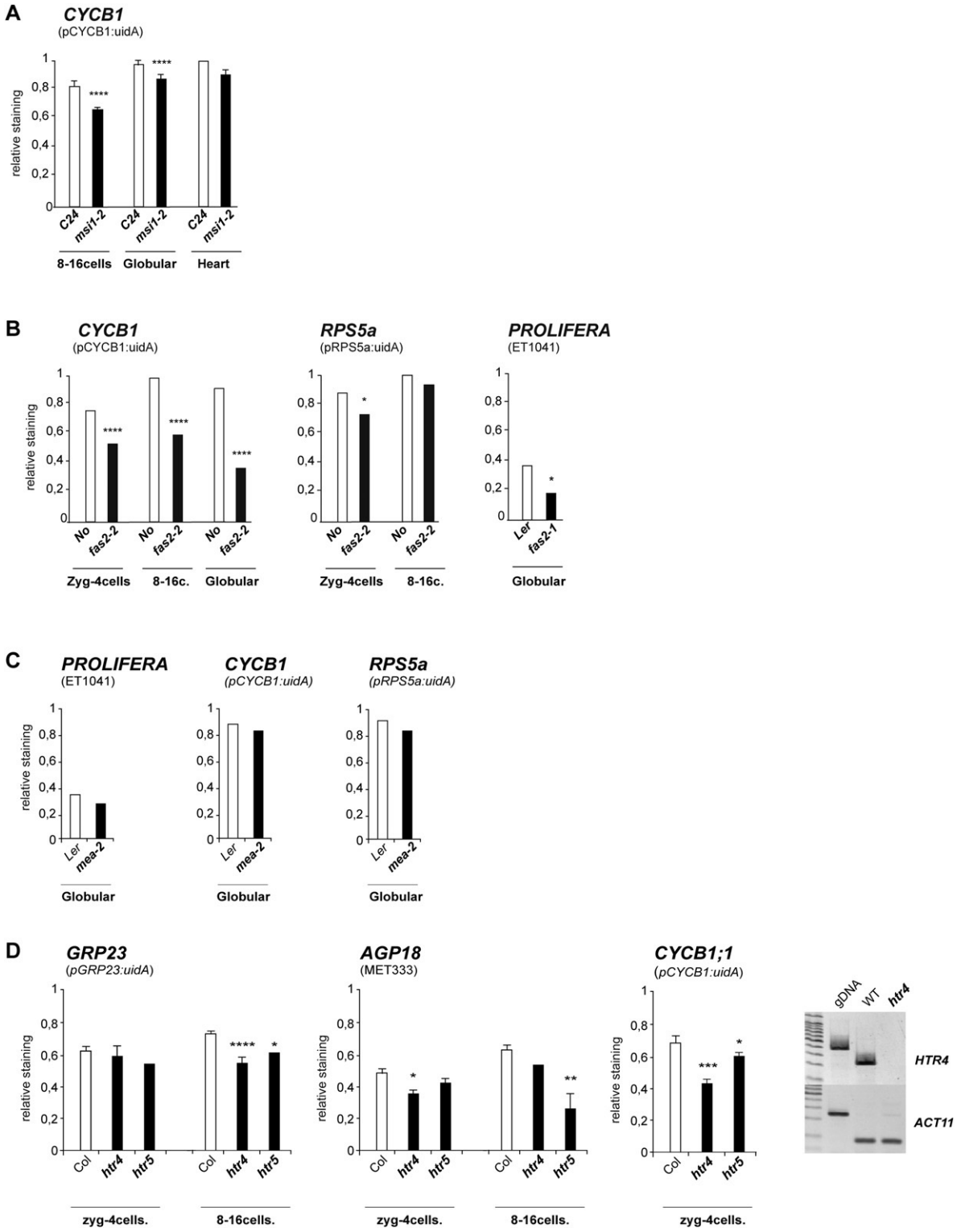


Figure S7. Maternal *MSI1* and *FAS2*, but not *MEA*, Control the Activation of Paternally Transmitted Markers and Require Maternal H3.3 Variants, Related to Figure 7

(A) Embryos showing GUS staining were scored in the progeny of crosses between wild-type (*Ler*), *msi1-2* females and the indicated paternal markers, at different developmental stages. Two-tailed Fisher's exact tests were carried out to assess the differences between the wild-type and mutant samples. The levels of significance are indicated (*: $p < 0,05$; **: $p < 0,01$; ***: $p < 0,001$; ****: $p < 0,0001$). Error bars represent SE. Detailed analysis is provided in Table S4.

(B) Embryos showing GUS staining were scored in the progeny of crosses between wild-type (No) or *fas2* females and the indicated paternal markers, at different developmental stages. Quantifications displayed as above. Detailed analysis is provided in Table S4.
(C) *MEA* does not delay expression of paternally transmitted markers at the globular stage. Embryos showing GUS staining were scored in the progeny of crosses between wild-type (L*er*) or *mea-2*/ *MEA* females and the indicated paternal markers, at the globular stage. Two-tailed Fisher's exact tests showed no significant differences. Detailed analysis is provided in Table S4.
(D) Maternal histone H3.3 variants control the activation of paternal markers. The graphs show embryos scored for GUS staining following crosses between wild-type (Col) or mutant females and the indicated paternal markers. Two-tailed Fisher's exact tests were carried out to assess the differences between wild-type and mutant samples as in A. Error bars represent SE. *htr4* and *htr5* are loss of function mutants in the *HTR4* and *HTR5* genes, which are close homologs encoding a H3.3 variant and likely show functional redundancy (Okada et al., 2005, Table S3). Absence of *HTR4* (*At4g40030*) mRNA in *htr4-1* insertion allele was confirmed by RT-PCR (see gel picture) performed on young siliques from wild-type (WT) and *htr4-1* (*htr4*) plants using intron-spanning primers. *ACTIN11* (*ACT11*) was amplified as a loading control. Amplification on genomic DNA (gDNA) is shown in the gel picture to the left.

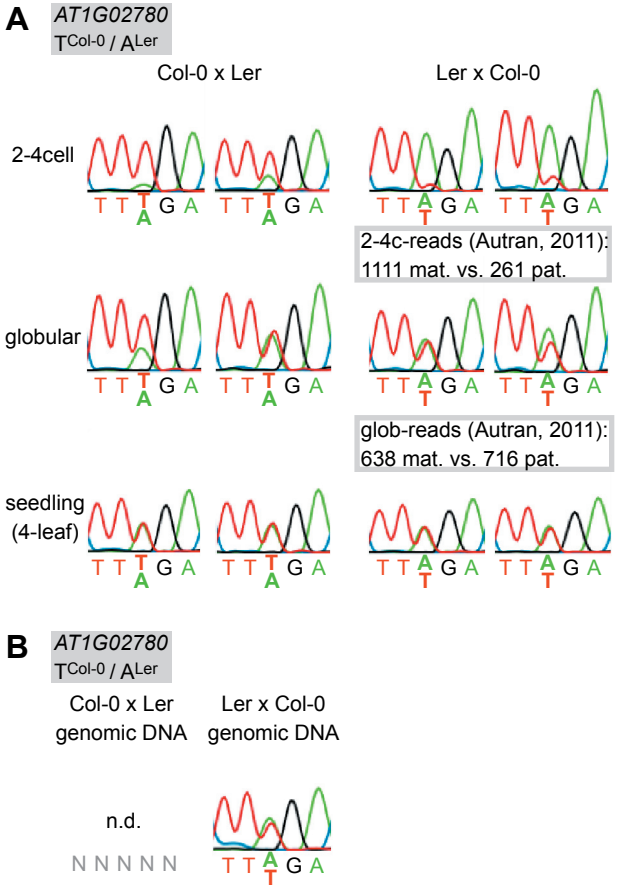


Figure S2-3-1. Allele-specific expression analysis of the biallelically expressed control gene *AT1G02780*. (A) Reciprocal hybrid embryos were isolated at 2.5 DAP (2-4 cell embryos) and at 4 DAP (globular embryos) and allele-specific expression was analyzed by RT-PCR and Sanger sequencing. The direction of the cross is indicated on top of each panel, the embryonic stage on the left. Two replicates were analyzed for each stage and cross, which is represented by two individual sequencing chromatograms. The analyzed gene and the polymorphism between Col-0 and Ler are indicated in the grey box. Furthermore, the SNP is displayed in bold below each chromatogram. The sequenced reads from (Autran et al., 2011) for AT1G02780 in Ler x Col-0 2-4 cell and globular embryo libraries are indicated below the chromatograms on the right hand side. In addition, the allele-specific expression was assessed on F1 hybrid seedling cDNA libraries (8 days after sowing). (B) Allele-specific PCR was performed on genomic DNA extracted from hybrid F1 seedlings in order to test whether the assay amplifies both alleles with equal efficiency.

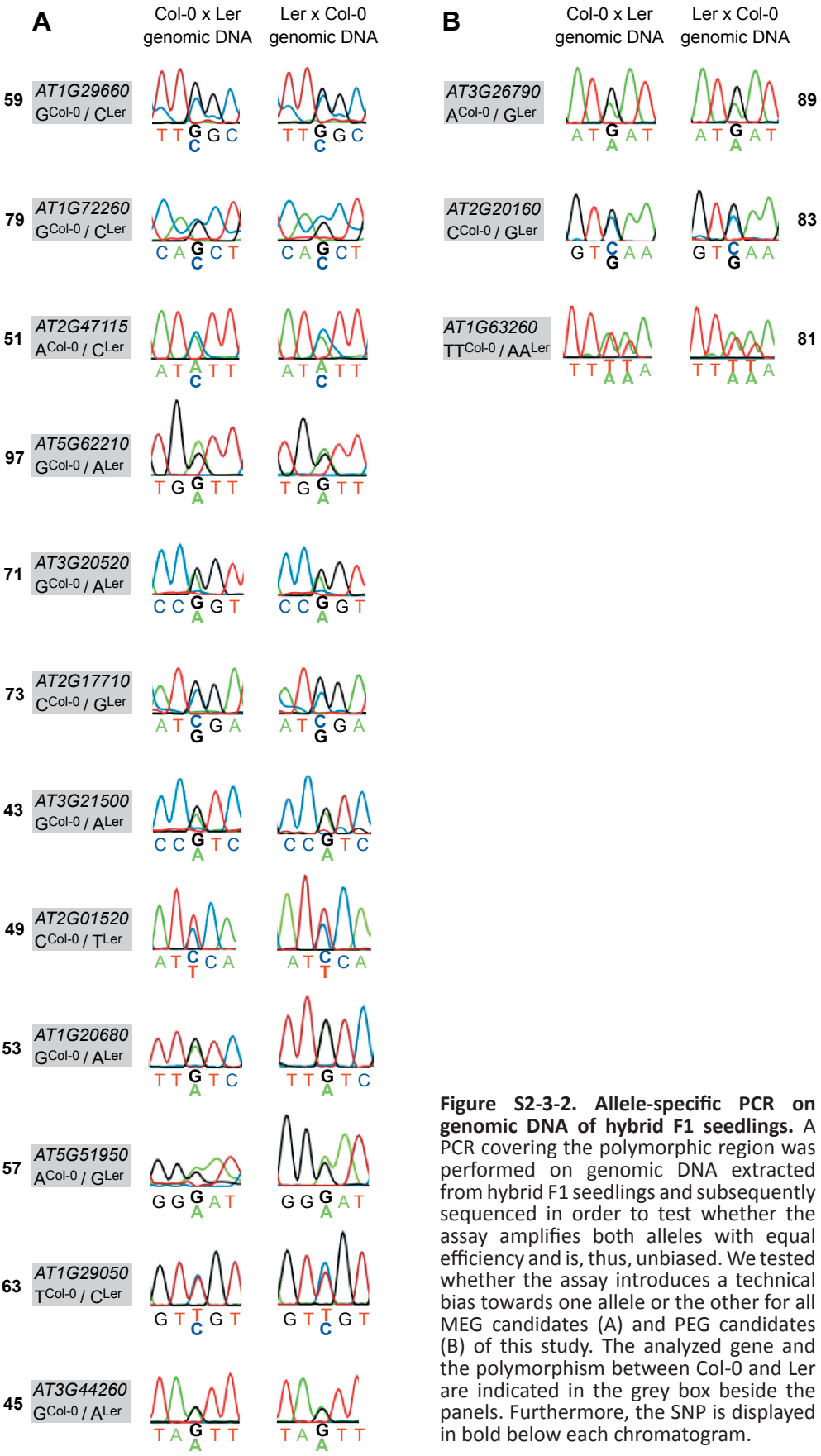


Figure S2-3-2. Allele-specific PCR on genomic DNA of hybrid F1 seedlings. A PCR covering the polymorphic region was performed on genomic DNA extracted from hybrid F1 seedlings and subsequently sequenced in order to test whether the assay amplifies both alleles with equal efficiency and is, thus, unbiased. We tested whether the assay introduces a technical bias towards one allele or the other for all MEG candidates (A) and PEG candidates (B) of this study. The analyzed gene and the polymorphism between Col-0 and Ler are indicated in the grey box beside the panels. Furthermore, the SNP is displayed in bold below each chromatogram.

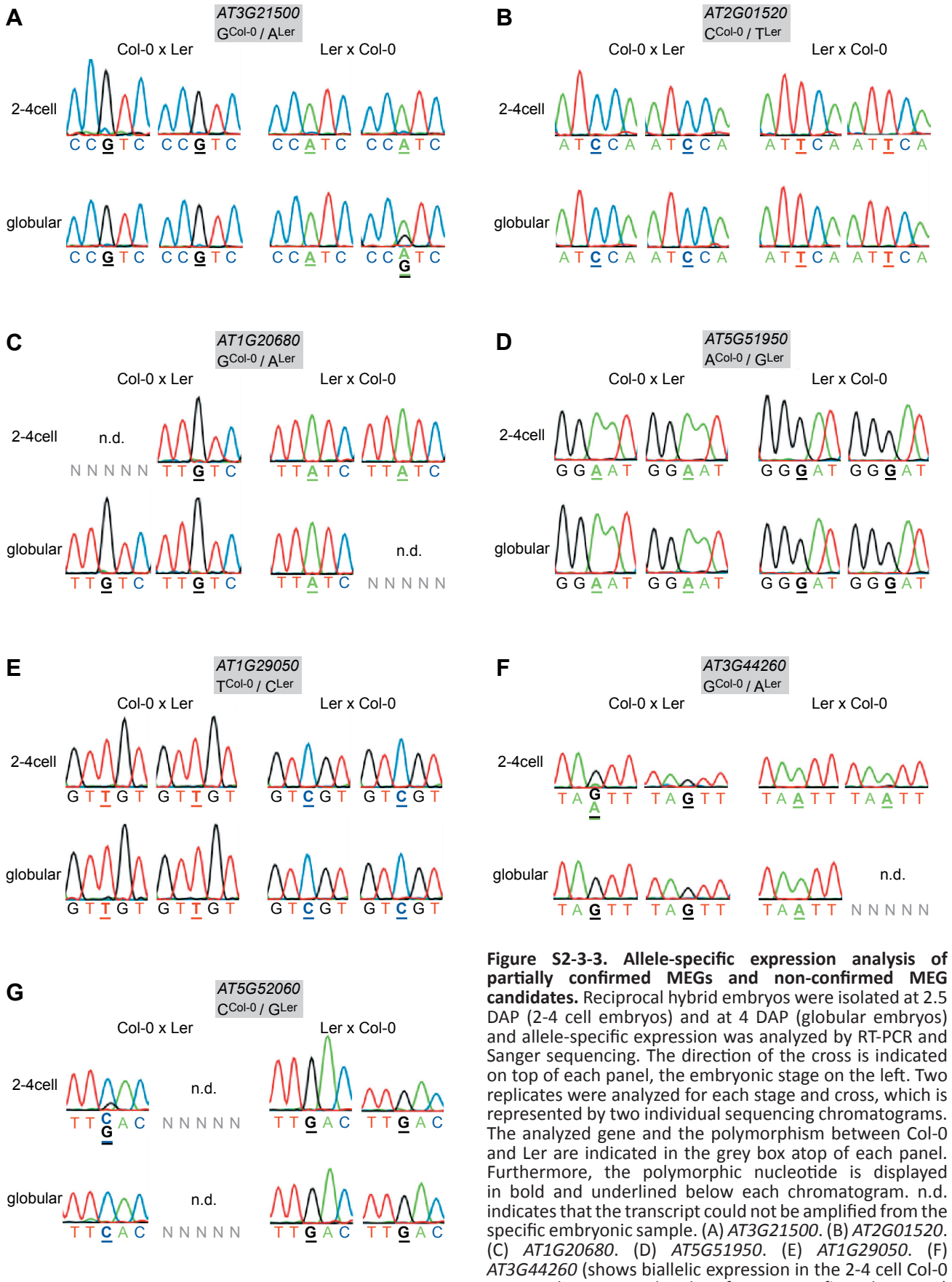


Figure S2-3-3. Allele-specific expression analysis of partially confirmed MEGs and non-confirmed MEG candidates. Reciprocal hybrid embryos were isolated at 2.5 DAP (2-4 cell embryos) and at 4 DAP (globular embryos) and allele-specific expression was analyzed by RT-PCR and Sanger sequencing. The direction of the cross is indicated on top of each panel, the embryonic stage on the left. Two replicates were analyzed for each stage and cross, which is represented by two individual sequencing chromatograms. The analyzed gene and the polymorphism between Col-0 and Ler are indicated in the grey box atop of each panel. Furthermore, the polymorphic nucleotide is displayed in bold and underlined below each chromatogram. n.d. indicates that the transcript could not be amplified from the specific embryonic sample. (A) *AT3G21500*. (B) *AT2G01520*. (C) *AT1G20680*. (D) *AT5G51950*. (E) *AT1G29050*. (F) *AT3G44260* (shows biallelic expression in the 2-4 cell Col-0 x Ler replicate #1 and is therefore not confirmed as MEG) (G) *AT5G52060* (shows biallelic expression in the 2-4 cell Col-0 x Ler replicate #1 and is therefore not confirmed as MEG).

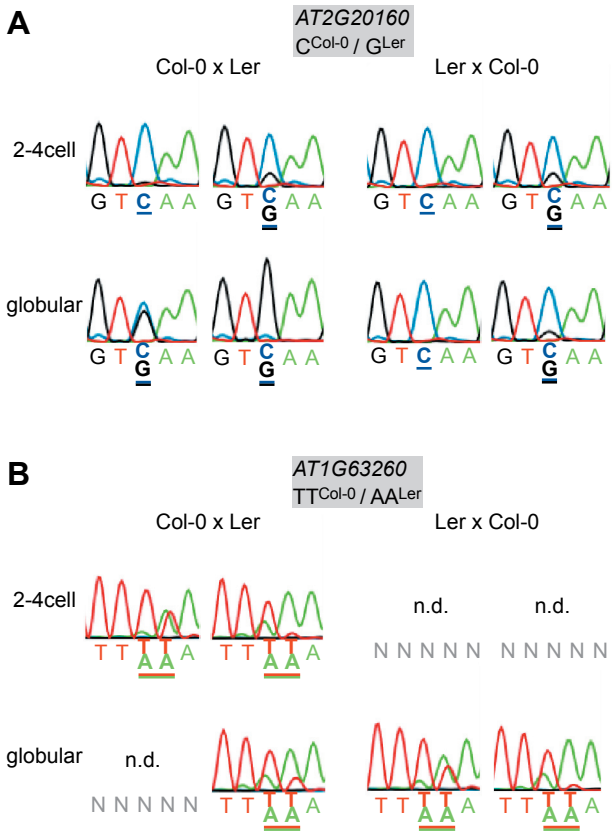


Figure S2-3-4. Allele-specific expression analysis of non-confirmed PEG candidates. Reciprocal hybrid embryos were isolated at 2.5 DAP (2-4 cell embryos) and at 4 DAP (globular embryos) and allele-specific expression was analyzed by RT-PCR and Sanger sequencing. The direction of the cross is indicated on top of each panel, the embryonic stage on the left. Two replicates were analyzed for each stage and cross, which is represented by two individual sequencing chromatograms. The analyzed gene and the polymorphism between Col-0 and Ler are indicated in the grey box atop of each panel. Furthermore, the polymorphic nucleotide is displayed in bold and underlined below each chromatogram. n.d. indicates that the transcript could not be amplified from the specific embryonic sample. (A) *AT2G20160* shows biallelic expression in 5 out of 8 samples. (B) *AT1G63260* shows biallelic expression in all samples from which the transcript was amplified.

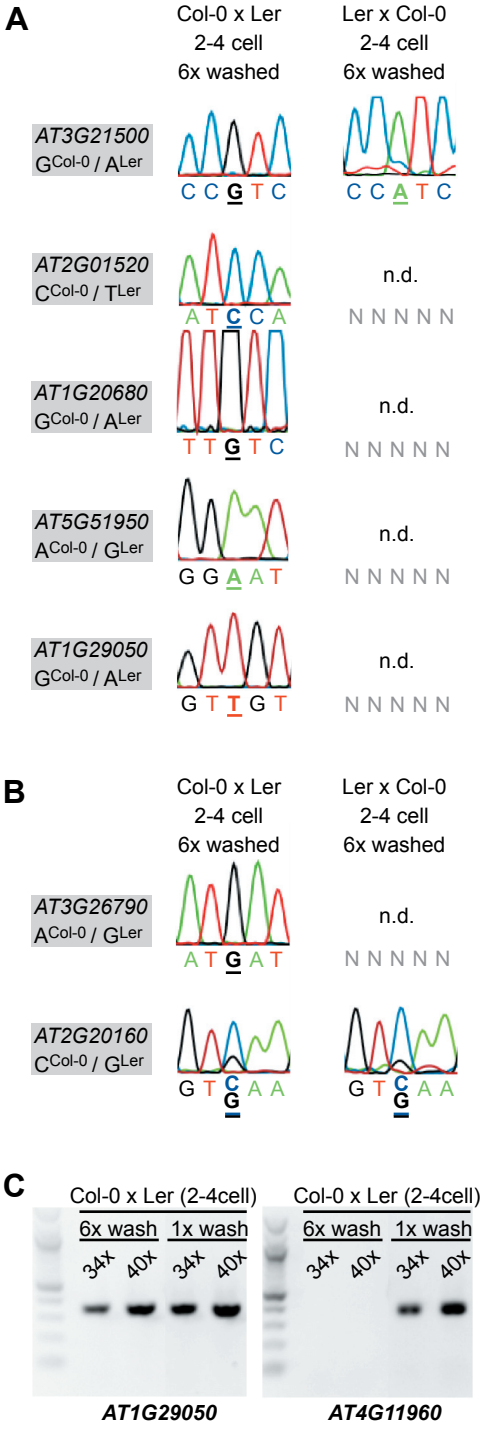


Figure S2-3-5. Allele-specific expression analysis of extensively washed embryonic control samples. Allele-specific expression analysis of confirmed or partially confirmed MEGs (A) and confirmed (*AT3G26790*) and non-confirmed (*AT2G20160*) PEGs (B) in the 6x washed embryonic samples. The analyzed gene and the polymorphism between Col-0 and Ler are indicated in the grey box. The polymorphic nucleotide is displayed in bold and underlined below each chromatogram. The transcript of some genes could not be amplified from the 6x washed 2-4 cell Ler x Col-0 sample likely due to library quality issues (indicated by n.d.). (C) Agarose gel analysis of the RT-PCR product of the partially confirmed MEG *AT1G29050* (left panel) and the non-confirmed *AT4G11960* (right panel).

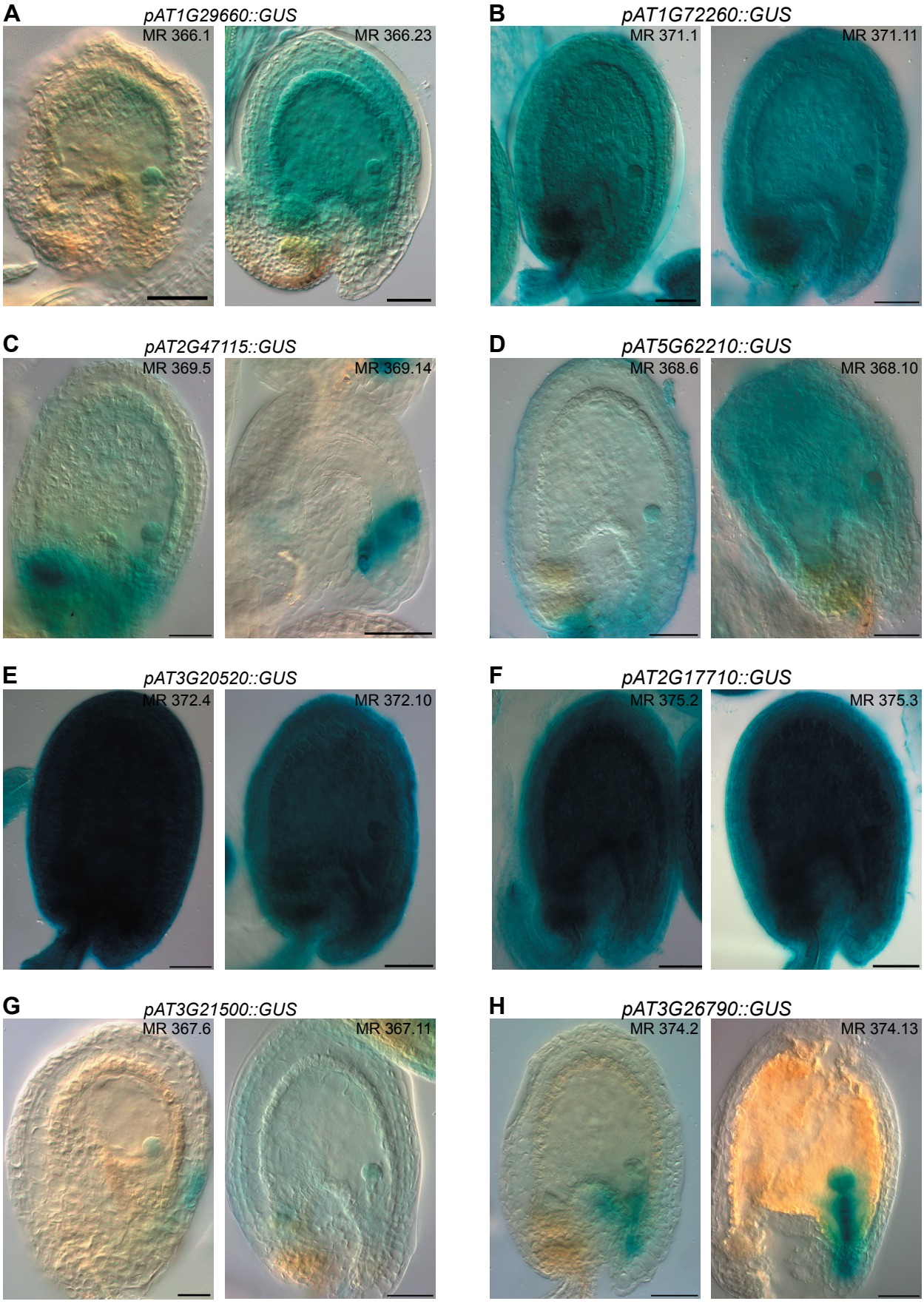


Figure S2-3-6. MEG and PEG reporter line analysis on whole seeds. Whole siliques were stained for GUS expression over night and analyzed for GUS signals in the seed and embryo. Almost all MEG reporter lines show a more or less strong expression in the seed coat (A-G). Yet, the PEG reporter FUS3::GUS is specifically expressed in the embryo (H). Each panel depicts the 2 strongest T1 lines that were used for further analysis. The reporter line is indicated on top of each panel and the individual line number in the upper right corner of each picture. Scale bar = 50 µm. (A) p*AT1G29660::GUS*. (B) *pAT1G72260::GUS*. (C) *pAT2G47115::GUS*. (D) *pAT5G62210::GUS*. (E) *pAT3G20520::GUS*. (F) *pAT2G17710::GUS*. (G) *pAT3G21500::GUS*. (H) *pAT3G26790::GUS* (*pFUS3::GUS*)

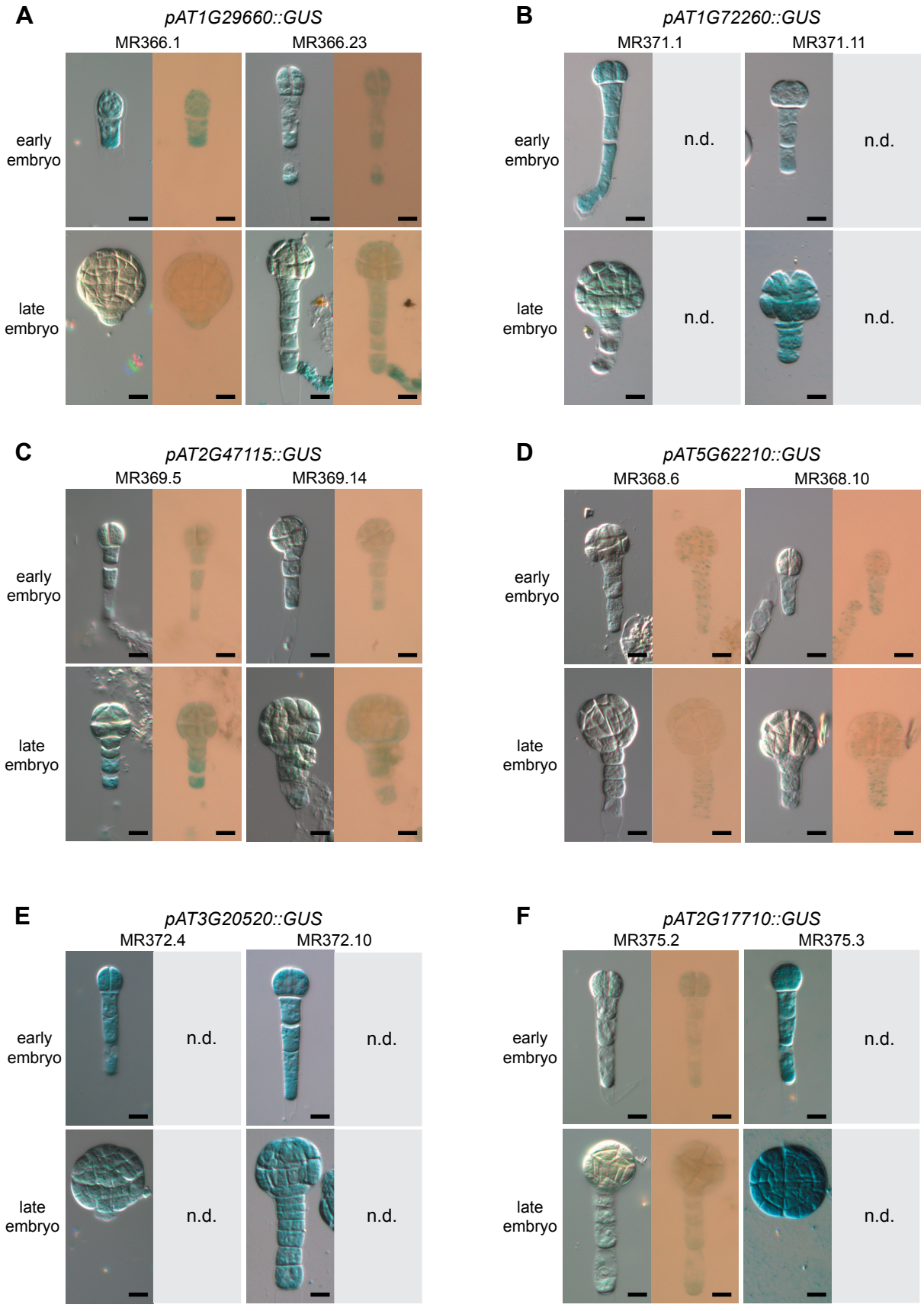


Figure S2-3-7. Embryo-specific expression of MEG reporter lines in isolated, self-fertilized embryos. Due to expression of all MEG reporters in the seed coat we isolated self-fertilized embryos carrying a MEG reporter line prior to GUS staining. Isolated embryos were released in GUS staining solution directly on a microscopic slide and were stained for 2-4 days at 37°C. 6 MEG reporter lines show a more or less strong and specific signal in the embryo. Each panel depicts two embryonic stages (early and late, indicated on the left), two independent lines (indicated on top) and a picture taken using DIC and bright-field microscopy (if not indicated otherwise). If no bright-field picture is shown (indicated by n.d.), then the signal was sufficiently visible when using DIC microscopy. Scale bar = 10 µm (A) p*AT1G29660::GUS*. (B) *pAT1G72260::GUS*. (C) *pAT2G47115::GUS*. (D) *pAT5G62210::GUS*. (E) *pAT3G20520::GUS*. (F) *pAT2G17710::GUS*.

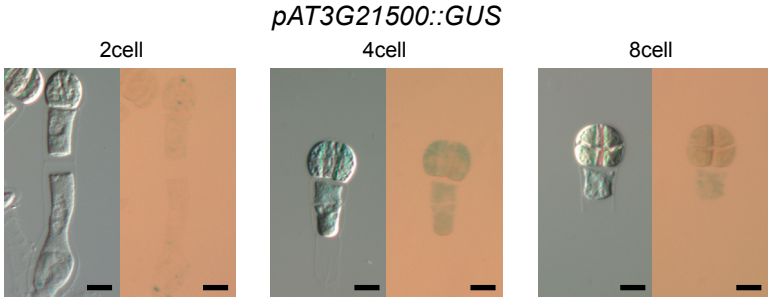


Figure S2-3-8. Embryo-specific expression of *pAT3G21500::GUS*. *pAT3G21500::GUS* was the weakest line in terms of embryonic expression. The reporter line is expressed in self-fertilized 4 cell and 8 cell embryos but shows no or very weak expression only in earlier stages. This line was not included in the parent-of-origin-dependent reporter expression analysis. Scale bar = 10 µm.

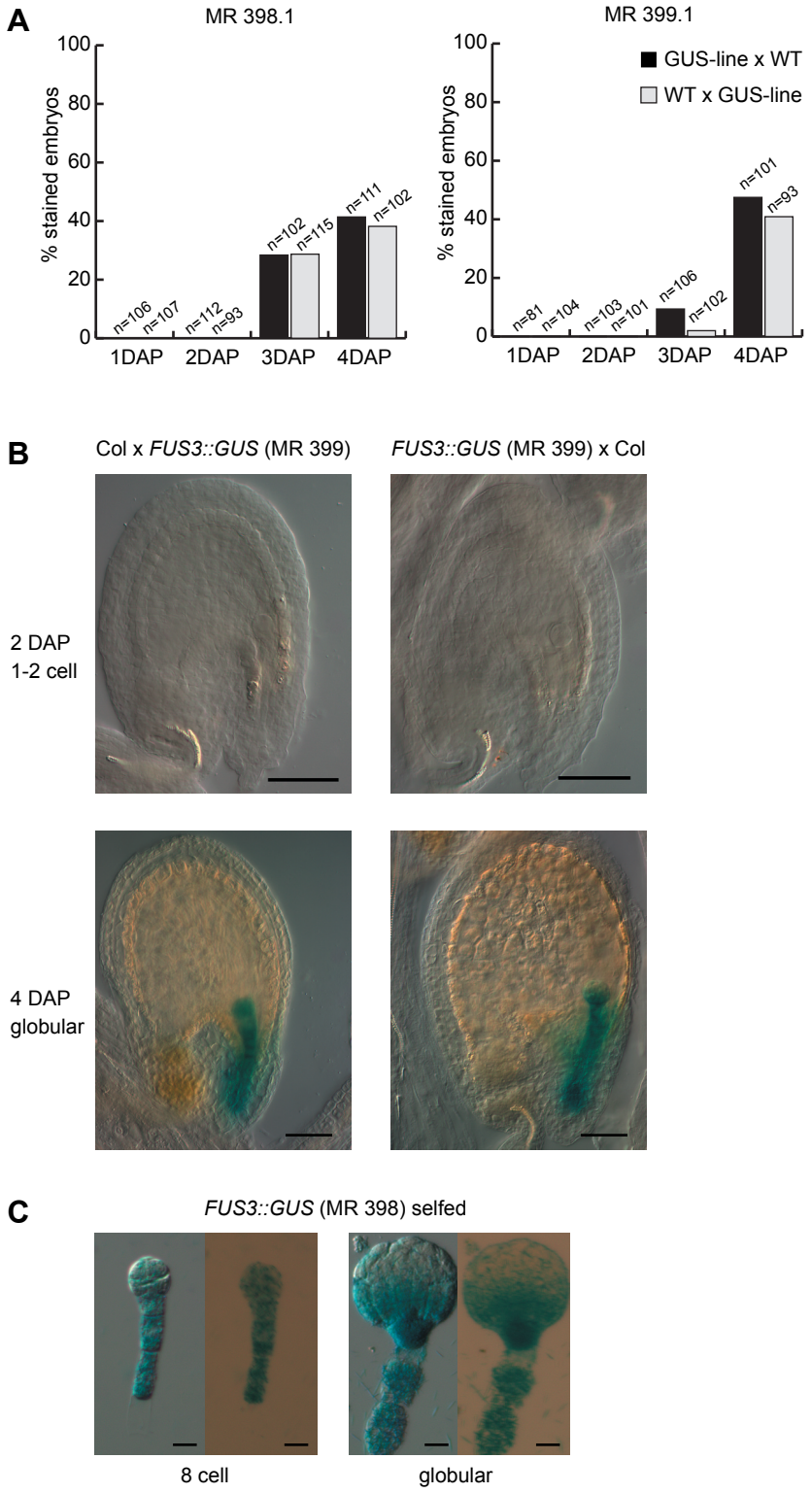


Figure S2-3-9. Parent-of-origin-dependent expression analysis of the PEG reporter line *pFUS3::GUS* (*pAT3G26790::GUS*). (A) Quantifications of reciprocal crosses of two independent insertions of *pFUS3::GUS* (MR 398 and MR 399). First signals were detected 3 DAP, coinciding with the first signal in isolated (4-)8 cell stage embryos. At 3 DAP the reporter is expressed from both parents already. GUS signal of the maternally inherited reporter is depicted in black, whereas GUS signal of the paternally inherited reporter is in grey. Numbers of counted seeds are indicated above each column. (B) Reciprocally crossed and stained seeds are shown 2 DAP (upper row) and 4 DAP (lower row). Whereas no GUS signal can be detected at 2 DAP, the reporter is clearly expressed from both parents 4 DAP. Scale bar = 50 µm. (C) Self-fertilized embryos were isolated at different timepoints and were stained on slide. Expression of the reporter was first detected at the (4-)8 cell stage. Scale bar = 10 µm.

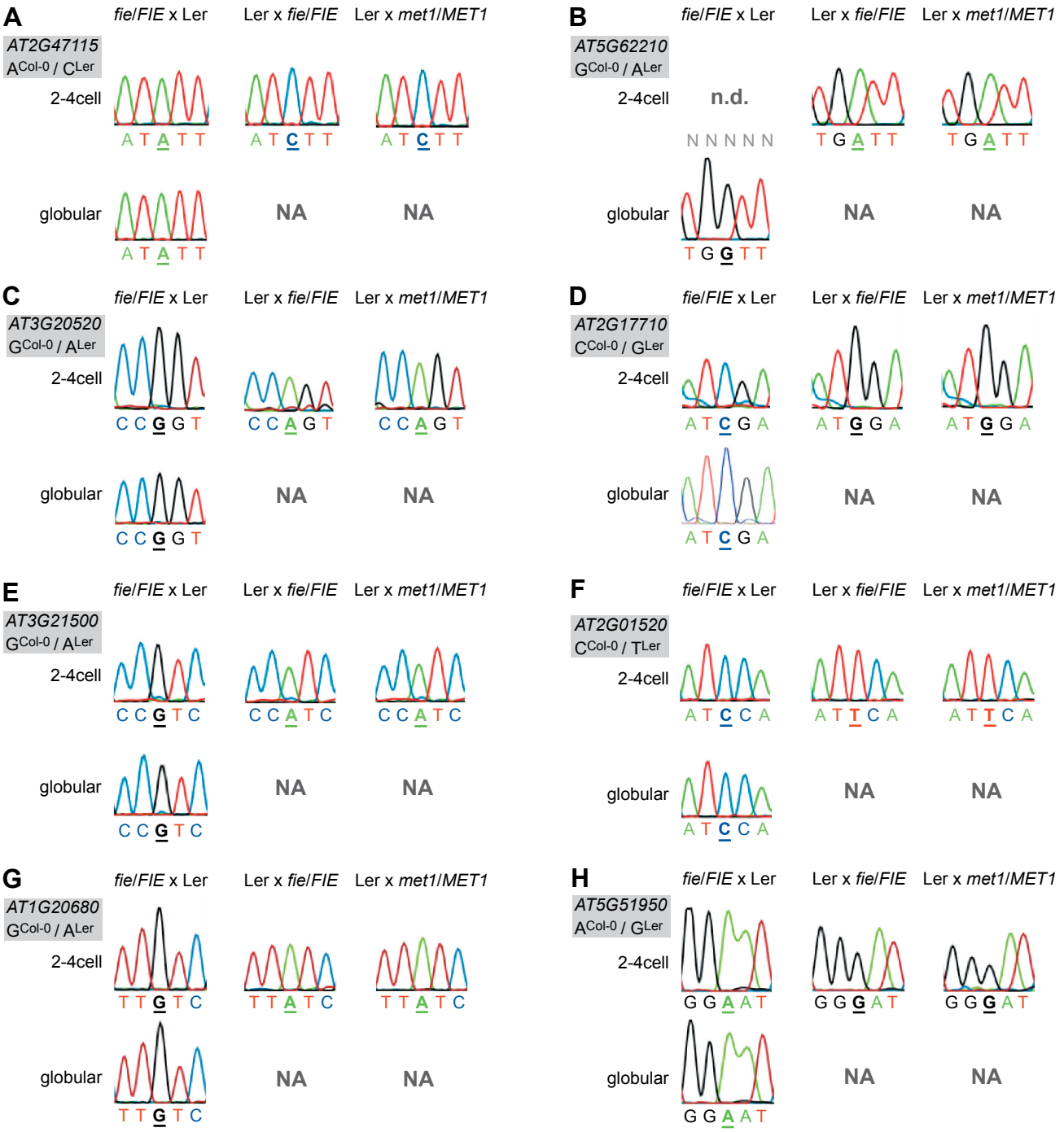


Figure S2-3-10. Effect of PRC2 and *MET1* function on the imprinted expression in the embryo. Mutant embryonic samples were generated and the confirmed MEGs (A-E) and partially confirmed MEGs (F-J) were analyzed for derepression of the silent allele. Heterozygous *fie* mutants were crossed maternally and paternally and heterozygous *met1-3* mutants were crossed paternally as indicated above the chromatograms. Embryos were isolated at 2.5 DAP (2-4 cell embryos) and at 4 DAP (globular embryos, only for the cross *fie*/*FIE* x Ler). The embryonic stage is indicated on the left, the analyzed gene and the polymorphism between the mutant (all in Col-0 background) and the wild-type allele (Ler) is shown in the grey box beside each panel. Furthermore, the polymorphic nucleotide is displayed in bold and underlined below each chromatogram. NA indicates that the library was not available. (A) *AT2G47115*. (B) *AT5G62210*. (C) *AT3G20520*. (D) *AT2G17710*. (E) *AT3G21500*. (F) *AT2G01520*. (G) *AT1G20680*. (H) *AT5G51950*. (J) *AT1G29050*. (K) *AT2G20160*.

Table S2-3-1. All potentially imprinted candidate genes in the *Arabidopsis thaliana* embryo from (Autran et al., 2011)**.** Table S1a gives a list of all candidate MEGs and PEGs in the embryo called in this study. Table S1b gives an overview of which class of genes were called, tested and confirmed. The tables can be found under the following link:
http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1003862#s5

Table S2-3-2. Phenotyping of T-DNA lines disrupting confirmed MEGs or the PEG. No obvious mutant phenotype was observed for any of the lines analyzed.

T-DNA line	locus	insertion site	seed set[1]	embryo patterning[2]	comments
SAIL_641_B05	*AT1G29660*	1st exon	wild-type	wild-type	MEG
SALK_023064C	*AT1G29660*	5th exon	wild-type	wild-type	MEG
SALK_039929C	*AT1G72260*	promoter	wild-type	wild-type	MEG
GABI_054H02	*AT1G72260*	3rd exon	wild-type	wild-type	MEG
FLAG_382E11	*AT2G47115*	5th exon	wild-type	wild-type	MEG
SALK_038225	*AT5G62210*	2nd exon	wild-type	wild-type	MEG
GABI_243D08	*AT5G62210*	intron	wild-type	wild-type	MEG
SAIL_330_H06	*AT5G62210*	promoter	wild-type	wild-type	MEG
SALK_010358.23.30	*AT3G20520*	2nd exon	wild-type	wild-type	MEG
SALK_152374.35.55	*AT3G20520*	4th exon	wild-type	wild-type	MEG
SALK 067511	*AT2G17710*	promoter	wild-type	wild-type	MEG
SALK 067639	*AT2G17710*	promoter	wild-type	wild-type	MEG
SALK_112762	*AT3G21500*	4th exon	wild-type	wild-type	MEG
SALK_039162.21.10	*AT3G26790*	1st exon	wild-type	wild-type	PEG; late phenotype (Raz et al. 2001)[3]

[1] mature siliques of 12-24 genotyped individuals were opened and examined for aborted seeds or infertile ovules.
[2] siliques at different stages were cleared using Hoyer's solution and analyzed for patterning defects.
[3] Raz et al. 2001 find a prolonged cell division phase in late embryos (torpedo stage). We analyzed earlier stages only.

Table S2-3-3. Embryonic expression of confirmed and non-confirmed MEGs and PEGs (data from Xiang et al., 2011).

YES = if value is bigger than 777.6, it is considered to be expressed
NO = if value is smaller than 777.6 it is considered background expression or low-expressed

confirmed MEGs and PEGs	Zygote	expressed?	Quadrant	expressed?	Globular	expressed?	Heart	expressed?	Torpedo	expressed?	Bent	expressed?	Mature	expressed?
AT1G29660	1826.0	YES	1809.9	YES	4244.0	YES	1813.1	YES	863.0	YES	1054.8	YES	809.2	YES
AT1G72260	1754.9	YES	2206.9	YES	1514.8	YES	1051.3	YES	718.8	NO	767.2	NO	790.9	YES
AT2G47115	768.8	NO	748.6	NO	762.2	NO	851.0	YES	855.9	YES	859.1	YES	782.1	YES
AT5G62210	1483.1	YES	1626.0	YES	1211.7	YES	1180.7	YES	1209.7	YES	1062.7	YES	1186.4	YES
AT3G20520	715.5	NO	682.7	NO	726.4	NO	766.9	NO	807.9	YES	716.9	NO	638.8	NO
AT2G17710	6331.6	YES	8454.1	YES	8818.0	YES	7114.0	YES	5394.2	YES	3736.1	YES	2025.5	YES
AT3G21500	4529.4	YES	6991.6	YES	2622.1	YES	1976.5	YES	1129.2	YES	1100.9	YES	903.5	YES
AT2G01520	3423.6	YES	8114.7	YES	9012.6	YES	5400.9	YES	1526.4	YES	2321.8	YES	4992.5	YES
AT1G20680	741.8	NO	650.7	NO	590.2	NO	655.7	NO	707.6	NO	632.2	NO	666.9	NO
AT5G51950	1015.6	YES	1184.6	YES	1229.5	YES	1155.8	YES	1182.1	YES	1030.6	YES	913.1	YES
AT1G29050	3504.4	YES	6785.6	YES	2309.3	YES	1551.6	YES	3352.1	YES	2795.3	YES	940.0	YES
AT3G26790	927.5	YES	1104.0	YES	1226.8	YES	1155.1	YES	1824.5	YES	4344.4	YES	1152.3	YES
AT1G02780	53776.0	YES	55938.5	YES	69276.8	YES	62158.1	YES	59432.6	YES	56060.6	YES	32089.4	YES

non-confirmed MEGs and PEGs	Zygote	expressed?	Quadrant	expressed?	Globular	expressed?	Heart	expressed?	Torpedo	expressed?	Bent	expressed?	Mature	expressed?
AT3G44260	756.1	NO	747.7	NO	750.8	NO	745.8	NO	767.2	NO	753.1	NO	803.5	YES
AT5G52060	1099.8	YES	1179.4	YES	970.9	YES	1030.3	YES	968.7	YES	928.2	YES	821.8	YES
AT2G20160	804.2	YES	838.4	YES	1011.1	YES	955.4	YES	867.5	YES	871.1	YES	786.0	YES
AT1G63260	2397.5	YES	2262.0	YES	3401.4	YES	3460.6	YES	2932.7	YES	2791.4	YES	3497.2	YES

Table S2-3-4. Parent-of-origin-dependent expression of confirmed MEGs and the PEG in other embryonic samples. One study analyzed torpedo-staged embryos in reciprocal crosses between Col-0 and Ler (Gehring et al., 2011), whereas the second study analyzed early embryo stages (1-2 cell, 8 cell, 32 cell) from reciprocal crosses between Col-0 and Cvi (Nodine and Bartel, 2012).

	Gehring et al. 2011. PLoS ONE 6: e23687						Nodine and Bartel. 2012. Nature 482: 94-97													
	torpedo stage						1-2cell				8cell				32cell					
	Col-0 x Ler		Ler x Col-0				Col-0 x Cvi		Cvi x Col-0		Col-0 x Cvi		Cvi x Col-0		Col-0 x Cvi		Cvi x Col-0			
	Reads overlapping SNPs		Reads overlapping SNPs				Reads overlapping SNPs		Reads overlapping SNPs		Reads overlapping SNPs		Reads overlapping SNPs		Reads overlapping SNPs		Reads overlapping SNPs			
Gene ID	Col reads	Ler reads	Col reads	Ler reads	parent-of-origin	comments	Col reads	Ler reads	Col reads	Ler reads	Col reads	Ler reads	Col reads	Ler reads	Col reads	Ler reads	Col reads	Ler reads	parent-pf-origin	comments
AT1G29660	0	2	0	1	Ler bias	few reads	10	0	0	0	1	3	0	1	0	0	6	6	bi-allelic	few reads
AT1G72260	0	0	0	2	maternal	few reads	0	0	0	4	2	0	0	1	0	0	0	0	maternal	few reads
AT2G47115	0	0	0	0	not expressed		0	0	0	0	0	0	0	5	0	0	0	6	maternal	few reads
AT5G62210	3	8	2	10	bi-allelic		0	0	0	0	0	0	0	11	4	0	0	5	maternal	
AT3G20520	1	0	1	3	maternal bias		0	0	0	14	0	0	0	9	2	0	0	5	maternal	
AT2G17710	19	3	7	44	maternal bias		23	7	0	13	11	0	4	10	9	0	9	15	bi-allelic	
AT3G21500	2	0	0	2	maternal	few reads	5	0	0	0	7	0	0	19	1	0	0	2	maternal	
AT2G01520	3	0	0	1	maternal	few reads	0	11	0	9	22	0	0	32	2	0	0	6	maternal bias	
AT1G20680	0	0	0	5	maternal	few reads	0	0	0	3	5	0	0	5	0	0	0	0	maternal	
AT5G51950	5	0	0	3	maternal	few reads	2	1	0	0	7	0	0	2	1	0	0	25	maternal bias	
AT1G29050	28	4	3	59	maternal bias		0	0	0	0	0	0	0	0	0	0	0	0	not expressed	
AT3G26790	112	75	80	63	bi-allelic		0	0	5	0	0	0	0	0	0	0	0	0	early PEG	
AT1G02780[1]	6111	4662	7734	6851	bi-allelic		0	1	0	0	1	0	0	1	0	1	0	0	not expressed	

[1] biallelic control gene

Table S2-3-5. All primers used in this study.

priMR number	sequence (5'-3')	general use	description
priMR 43	TCCAAATCCTCCAATGGAAC	RT-PCR & Sanger	AT3G21500
priMR 44	TGACCGTCCTTCTTGCTTTC	RT-PCR & Sanger	AT3G21500
priMR 45	AAGATCGTCACCGAGGTCTG	RT-PCR & Sanger	AT3G44260
priMR 46	TCGTGACTCGTGAAGTCTGG	RT-PCR & Sanger	AT3G44260
priMR 49	TTCACTCGCTACACAAGGAAGA	RT-PCR & Sanger	AT2G01520
priMR 50	TCGTTTTGCTTCTCCCAGAT	RT-PCR & Sanger	AT2G01520
priMR 51	TGGGAAACAGGAGTTCTTGG	RT-PCR & Sanger	AT2G47115
priMR 52	TGCTGACGGACATTGTGTTT	RT-PCR & Sanger	AT2G47115
priMR 53	GCAAAGCTAACAACGGAACC	RT-PCR & Sanger	AT1G20680
priMR 54	TGGCAGAATCTGTAGAATCGAG	RT-PCR & Sanger	AT1G20680
priMR 57	GCTCTCGTGGTAGCATTCAA	RT-PCR & Sanger	AT5G51950
priMR 58	CGATGGTAGGGCAAGGTATG	RT-PCR & Sanger	AT5G51950
priMR 59	TTCATGCCTCAGTTCTACTCCA	RT-PCR & Sanger	AT1G29660
priMR 60	GCGCCATAAGCATTGATGTA	RT-PCR & Sanger	AT1G29660
priMR 61	CCCAGGTTATTTCTGGGAAG	RT-PCR & Sanger	At1G24090 / At1G24095
priMR 62	GGCTCAGCTGCTTTAGACTG	RT-PCR & Sanger	At1G24090 / At1G24095
priMR 63	GCTGCGTAGAGAAGTTGGTTC	RT-PCR & Sanger	AT1G29050
priMR 64	GGGTTGATCAAAACGTCGAT	RT-PCR & Sanger	AT1G29050
priMR 65	AAAGAATATGCTCTTGTGTATTTGAA	RT-PCR & Sanger	AT5G42650
priMR 66	ATAAACAATGCGCCGGTAAG	RT-PCR & Sanger	AT5G42650
priMR 115	GTGGTTCAAACCGGAGACAT	RT-PCR & Sanger	AT5G42650
priMR 116	GGGTGGGATGAAGATTTTGTT	RT-PCR & Sanger	AT5G42650
priMR 67	CAAGATGGTAAGATTGGGATAGTG	RT-PCR & Sanger	AT2G44470
priMR 68	GGAGATGAGCGGTGAAGTGT	RT-PCR & Sanger	AT2G44470
priMR 71	TCATCCTTGAAAAGGAGACAAA	RT-PCR & Sanger	AT3G20520
priMR 72	GAGAAGAAGCAGGAGGCTGA	RT-PCR & Sanger	AT3G20520
priMR 117	AAGTGTTACGGCGAATTTGG	RT-PCR & Sanger	AT3G20520
priMR 118	TTTTCCTCTGCCTTTGATGG	RT-PCR & Sanger	AT3G20520
priMR 73	TCTCTTTAGTTTGAAGGTTAGATTCAG	RT-PCR & Sanger	AT2G17710
priMR 74	TGATCTTCGGTTTCCAGGTC	RT-PCR & Sanger	AT2G17710
priMR 137	CGCTTCTTCCTCCATCTCAG	RT-PCR & Sanger	AT2G17710
priMR 138	ACACGAGCTGCAGCCTTAAT	RT-PCR & Sanger	AT2G17710
priMR 77	AACATGTACCTGCGAGCAAA	RT-PCR & Sanger	AT3G58780
priMR 78	TTTGGATCTCGACGTGTTTT	RT-PCR & Sanger	AT3G58780
priMR 79	GAGTGTTCATGGCACCACAC	RT-PCR & Sanger	AT1G72260
priMR 80	TCCCTTGTGATCAAACAAGTAAA	RT-PCR & Sanger	AT1G72260
priMR 81	TGTTGTGGTTTGGCAATCTT	RT-PCR & Sanger	AT1G63260
priMR 82	GTTATCCTGCGGTGAATGCT	RT-PCR & Sanger	AT1G63260
priMR 103	CTGTTGTGGTTTGGCAATCTT	RT-PCR & Sanger	AT1G63260
priMR 104	GCGATCTTCAATGTGGTCCT	RT-PCR & Sanger	AT1G63260
priMR 168	GGAGAATCCGGTGAAGTGTT	RT-PCR & Sanger	AT1G63260
priMR 169	GCCAAACCACAACAGTACCA	RT-PCR & Sanger	AT1G63260
priMR 83	TGCGAAGCTCTTCTTCTTCC	RT-PCR & Sanger	AT2G20160
priMR 84	CAGATGGTAGCGCACATGAT	RT-PCR & Sanger	AT2G20160
priMR 135	TGCGAAGCTCTTCTTCTTCC	RT-PCR & Sanger	AT2G20160
priMR 136	CGCAGAGTTCGTGAAAAACA	RT-PCR & Sanger	AT2G20160
priMR 85	CACTGGCTCATCTCCACATTT	RT-PCR & Sanger	AT2G24430
priMR 86	CGACTTCACTCCAAGTCCTCA	RT-PCR & Sanger	AT2G24430
priMR 105	AAGTCCATGCACTGGCTCAT	RT-PCR & Sanger	AT2G24430
priMR 106	AGCACCACCATGGACTCTTC	RT-PCR & Sanger	AT2G24430
priMR 87	AAGTGCTTGGTGAGCAACAG	RT-PCR & Sanger	AT3G63480
priMR 88	AGTAGAAGCCACTGCGCCTA	RT-PCR & Sanger	AT3G63480
priMR 107	GCATCCTGTAAGATGCGAGT	RT-PCR & Sanger	AT3G63480
priMR 108	TGGCTCAGAGAAAGCAGACA	RT-PCR & Sanger	AT3G63480
priMR 89	CATACGTCGGAGAGAGCTGA	RT-PCR & Sanger	AT3G26790
priMR 90	AAACCCAAAGAGATCCACCA	RT-PCR & Sanger	AT3G26790
priMR 109	CACCATCAAGAATACCATGAGC	RT-PCR & Sanger	AT3G26790
priMR 110	GACACGACGGCTCTTACCTC	RT-PCR & Sanger	AT3G26790
priMR 93	CTTTCCTAATGAGAGCAAAGTTTC	RT-PCR & Sanger	AT4G11960
priMR 94	TCCCCAATGATAGTAACACCAA	RT-PCR & Sanger	AT4G11960
priMR 95	CATGCTTCCTTTCTCGGCTA	RT-PCR & Sanger	AT1G02780 (BiG)
priMR 96	ATGAAGGGTCGTCACTCTGG	RT-PCR & Sanger	AT1G02780 (BiG)
priMR 97	ACGCAACAACCAAACTCTGA	RT-PCR & Sanger	AT5G62210
priMR 98	GAAGGTTCCAAATCCGTGAA	RT-PCR & Sanger	AT5G62210
priMR 99	CTCCAACCAAAAGACCTTGG	RT-PCR & Sanger	AT1G57800
priMR 100	CCCCGTATCAACTTGTCCAT	RT-PCR & Sanger	AT1G57800
priMR 113	TCTTGCCACTGGAAGCATTT	RT-PCR & Sanger	AT1G57800
priMR 114	CCCCGTATCAACTTGTCCAT	RT-PCR & Sanger	AT1G57800
priMR 101	GGCTCCTCTTCAATCTCTTGC	RT-PCR & Sanger	AT5G52060
priMR 102	TGTCAAGTTACAAAGAAAGATGCAG	RT-PCR & Sanger	AT5G52060
priMR 119	TCAAAGTTCTGCAGATTATCCG	genotyping	SALK_112762.41.75 = 43-1
priMR 120	TTCCACAATTAGATTGCAGCC	genotyping	SALK_112762.41.75 = 43-1
priMR 121	TAGATCGAACAACCGTGGTTC	genotyping	FLAG_382E11 = 51-1
priMR 122	GGGATATGAACAATTGCCATG	genotyping	FLAG_382E11 = 51-1
priMR 123	GAAAACACAGCTGCCTAGCTG	genotyping	SAIL_641_B05 = 59-1
priMR 124	CTTCTCTGATTCCAGCAGCTG	genotyping	SAIL_641_B05 = 59-1
priMR 125	ATTCGATGGAAAACTATGGGG	genotyping	SALK_023064.22.75 = 59-2
priMR 126	TAGATCCAGTTTCAGCGAAGC	genotyping	SALK_023064.22.75 = 59-2
priMR 127	ACGCTCCATTCACAATTTCAC	genotyping	SALK_039929.55.25 = 79-1
priMR 128	GAACTCTATGCGTGCCAAGAG	genotyping	SALK_039929.55.25 = 79-1
priMR 129	TGTATTTTCATGGTGTGTGGTTG	genotyping	SALK_038225.46.35 = 97-1
priMR 130	TCCTTCTCTAACCTTCGAGCC	genotyping	SALK_038225.46.35 = 97-1
priMR 131	TGTATTTTCATGGTGTGTGGTTG	genotyping	GABI_243D08 = 97-2
priMR 132	TCCTTCTCTAACCTTCGAGCC	genotyping	GABI_243D08 = 97-2
priMR 133	TGAGCTAATTTGGTAGGTGGG	genotyping	SAIL_330_H06 = 97-3
priMR 134	TAATGGCTTTGGTTTATCCGG	genotyping	SAIL_330_H06 = 97-3
priMR 139	tctgcagattatccgtgagaaa	genotyping	SALK_112762.41.75 = 43-1 **(replacing priMR 119)**

Table S2-3-5 continued. All primers used in this study.

priMR 140	ACCCTAACTTGCAGTGCTCA	genotyping	SALK_039929.55.25 = 79-1 **(replacing priMR 127)**
priMR 165	CAAAGGAACACTCTCATTCGC	genotyping	SALK_067511 & SALK_067639 = 73-1 & 73-2
priMR 166	TCCAACGTCTACGTCTACCTTTC	genotyping	SALK_067511 & SALK_067639 = 73-1 & 73-3
priMR 170	GAAACCAGACTCAAGCTGTGC	genotyping	SALK_010358.23.30 = 71-1
priMR 171	TTTGCGTAGAAAGCACTGTCC	genotyping	SALK_010358.23.30 = 71-1
priMR 172	GGCTGCAATAAAGAGGGAAAC	genotyping	SALK_152374.35.55 = 71-2
priMR 173	AGCAAAGTTGACAGCTTCAGG	genotyping	SALK_152374.35.55 = 71-2
priMR 174	TGGTTAGCCGTTAAAACGAAG	genotyping	GABI_054H02 = 79-2
priMR 175	TGTTGCTAAGTCGCATCTGTG	genotyping	GABI_054H02 = 79-2
priMR 180	TGGAAACTTTTTATTGCGGG	genotyping	SALK_034196.49.40 = 83-1
priMR 181	GCGCTACCATCTGCATCTTAC	genotyping	SALK_034196.49.40 = 83-1
priMR 182	GTAAGATGCAGATGGTAGCGC	genotyping	FLAG_148A09 = 83-2
priMR 183	TCTGCAGGTTTGACCATTACC	genotyping	FLAG_148A09 = 83-2
priMR 184	AACCGTAAACCCGACATCTTC	genotyping	SALK_039162.21.10 = 89-1
priMR 185	TACCAGAAACTCCATCAACGG	genotyping	SALK_039162.21.10 = 89-1
priMR 141	aaaaagcaggctTCCAAGCAATGAGCGAAA	cloning	LP + attB1 for **AT1G29660** amplifying 2461bp promoter
priMR 142	agaaagctgggtCGATCTCACAAAACAAAAC	cloning	RP + attB2 for **AT1G29660** amplifying 2461bp promoter
priMR 143	aaaaagcaggctGGGCCCGTACTAAACATCC	cloning	LP + attB1 for **AT3G21500** amplifying 1441bp promoter
priMR 144	agaaagctgggtCCAACGTATATATAGTGG	cloning	RP + attB2 for **AT3G21500** amplifying 1441bp promoter
priMR 145	aaaaagcaggctCAAATGGAGATGATGATTGGA	cloning	LP + attB1 for **AT5G62210** amplifying 991bp promoter
priMR 146	agaaagctgggtTGTTGTCTTCTTGAAAAGAG	cloning	RP + attB2 for **AT5G62210** amplifying 991bp promoter
priMR 147	aaaaagcaggcttcgggacttttttacattttgaa	cloning	LP + attB1 for **AT2G47115** amplifying 480bp promoter
priMR 167	agaaagctgggtCGATCAATGCCGTgattt	cloning	RP + attB2 for **AT2G47115** amplifying 480bp promoter
priMR 151	aaaaagcaggctgttggcctacggcacttg	cloning	LP + attB1 for **AT1G72260** amplifying 357bp promoter
priMR 152	agaaagctgggtcttttgaatggtttacttg	cloning	RP + attB2 for **AT1G72260** amplifying 357bp promoter
priMR 153	aaaaagcaggcttccgttgttgaatttgtcg	cloning	LP + attB1 for **AT3G20520** amplifying 826bp promoter
priMR 154	agaaagctgggtggtagtggagagctttca	cloning	RP + attB2 for **AT3G20520** amplifying 826bp promoter
priMR 157	aaaaagcaggcttggtatgcaatccctcca	cloning	LP + attB1 for **AT3G26790** amplifying 2220bp promoter
priMR 158	agaaagctgggtctctctcaattggttaac	cloning	RP + attB2 for **AT3G26790** amplifying 2220bp promoter
priMR 159	aaaaagcaggctacttgccatcggtcttgg	cloning	LP + attB1 for **AT2G17710** amplifying 858bp promoter
priMR 160	agaaagctgggtaccacaacttcacctttc	cloning	RP + attB2 for **AT2G17710** amplifying 858bp promoter
ACT11_F	AACTTTCAACACTCCTGCCATG	RT-PCR control gene	amplification of ACTIN 11 transcripts
ACT11_R	CTGCAAGGTCCAAACGCAGA	RT-PCR control gene	amplification of ACTIN 11 transcripts
WOX9_F	ccatcaacttcggaccagctt	RT-PCR control gene	amplification of WOX9 transcripts
WOX9_R	tccctcacattgaacggtcct	RT-PCR control gene	amplification of WOX9 transcripts
LBb1.3	ATTTTGCCGATTTCGGAAC	genotyping	genotyping SALK lines
GBF_AC161_LB1	ATATTGACCAT ATACTCATTGC	genotyping	genotyping GABI-Kat lines
FL_LB4	CGTGTGCCAGGTGCCCACGGAATAGT	genotyping	genotyping FLAG lines
Syg_LB1	GCCTTTTCAGAATGGATAAATAGCCTTGCTTCC	genotyping	genotyping SAIL lines

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Genetics & Development

Epigenetic regulation and reprogramming during gamete formation in plants

Célia Baroux, Michael T Raissig and Ueli Grossniklaus

Plants and animals reproduce sexually via specialized, highly differentiated gametes. Yet, gamete formation drastically differs between the two kingdoms. In flowering plants, the specification of cells destined to enter meiosis occurs late in development, gametic and accessory cells are usually derived from the same meiotic product, and two distinct female gametes involved in double fertilization differentiate. This poses fascinating questions in terms of gamete development and the associated epigenetic processes. Although studies in this area remain at their infancy, it becomes clear that large-scale epigenetic reprograming, involving RNA-directed DNA methylation, chromatin modifications, and nucleosome remodeling, contributes to the establishment of transcriptionally repressive or permissive epigenetic landscapes. Furthermore, a role for small RNAs in the regulation of transposable elements during gametogenesis is emerging.

Address
Institute of Plant Biology, Zürich-Basel Plant Science Center, University of Zürich, Zollikerstrasse 107, CH-8008 Zürich, Switzerland

Corresponding author: Baroux, Célia (cbaroux@botinst.uzh.ch)

Current Opinion in Genetics & Development 2011, **21**:124–133

This review comes from a themed issue on
Chromosomes and expression mechanisms
Edited by Peter Verrijzer and Ali Shilatifard

Available online 14th February 2011

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2011.01.017

Introduction

In multicellular organisms, sexual reproduction involves the union of highly specialized, haploid gametes. Between plants and animals the ontology of the gametes as well as the fate of the fertilization products differ [1,2,3•]. In particular, plants do not set aside a germ line lineage early in development as animals do. Instead spore mother cells (SMCs), the cells destined to undergo meiosis, are specified late during the development of the diploid generation of the plant life cycle, the sporophyte. In flowering plants this occurs in specialized male and female floral organs, the anthers and ovules, respectively. Moreover, the meiotic products in plants, the spores, do not differentiate directly into gametes as in animals; rather, they divide mitotically to produce multicellular, haploid gametophytes. These can be autonomous, free living organism as for instance in mosses and ferns [4] or highly reduced to a few cells as in flowering plants [5,6]. In the majority of flowering plants, the male gametophyte (pollen) consists of three cells: two gametic sperm cells harbored within an accessory, vegetative cell responsible to deliver the sperm cells to the female gametes. The female gametophyte (embryo sac) produces two gametes, the egg and central cell, and five accessory cells: two synergids assisting fertilization and three antipodals [5–7]. Importantly, double fertilization generates two fertilization products with distinct developmental fates: while the fertilized egg gives rise to the embryo, the fertilized central cell generates an extra-embryonic nurturing tissue, the endosperm [7]. This complex scheme of gamete formation has implications in terms of developmental strategies governing first, sporogenesis in sporophytic ('somatic') floral tissues, second the specification of gametic versus accessory fate in the gametophytes, and third the distinct developmental fates of the egg and central cells. Not surprisingly, a role for epigenetic processes in these different events has been recognized [3•,8] and is supported by recent studies [9••,10••,11•,12••] which we will discuss here.

Epigenetic patterns in sporogenic fate control and meiotic progression

In flowering plants, on which our review will focus, undifferentiated stem cells reside in apically localized meristems. These meristems produce cells contributing to vegetative (root, shoot) and floral tissues [13] in which SMCs will be specified. Yet, there is no germ line lineage with a traceable sporogenic fate [14]. Instead, SMCs are newly specified in a subepidermal position within multicellular sporangia, the ovule primordium and anther locule, respectively [1]. The highly regular, predictable position of SMCs suggests a specification process dependent on the cell's position [3•,15,16] rather than its lineage as in animals, a developmental strategy commonly used in plants [17]. Consistent with this idea, maize, rice and *Arabidopsis* mutants lacking specific Leu-rich repeat receptor-like protein kinases, potentially involved in cellular signaling, produce supernumerary SMCs, the microspore and/or megaspore mother cells, respectively [16,18,19]. The observed phenotypes suggest that subepidermal cells surrounding the SMCs share a sporogenic potential [3•,20•,21]. Recent findings indicate that the sporogenic fate is epigenetically suppressed in cells other than the SMCs. *Arabidopsis* mutants lacking *AGO9* function develop multiple megaspore mother cells (MMCs) in the nucellus

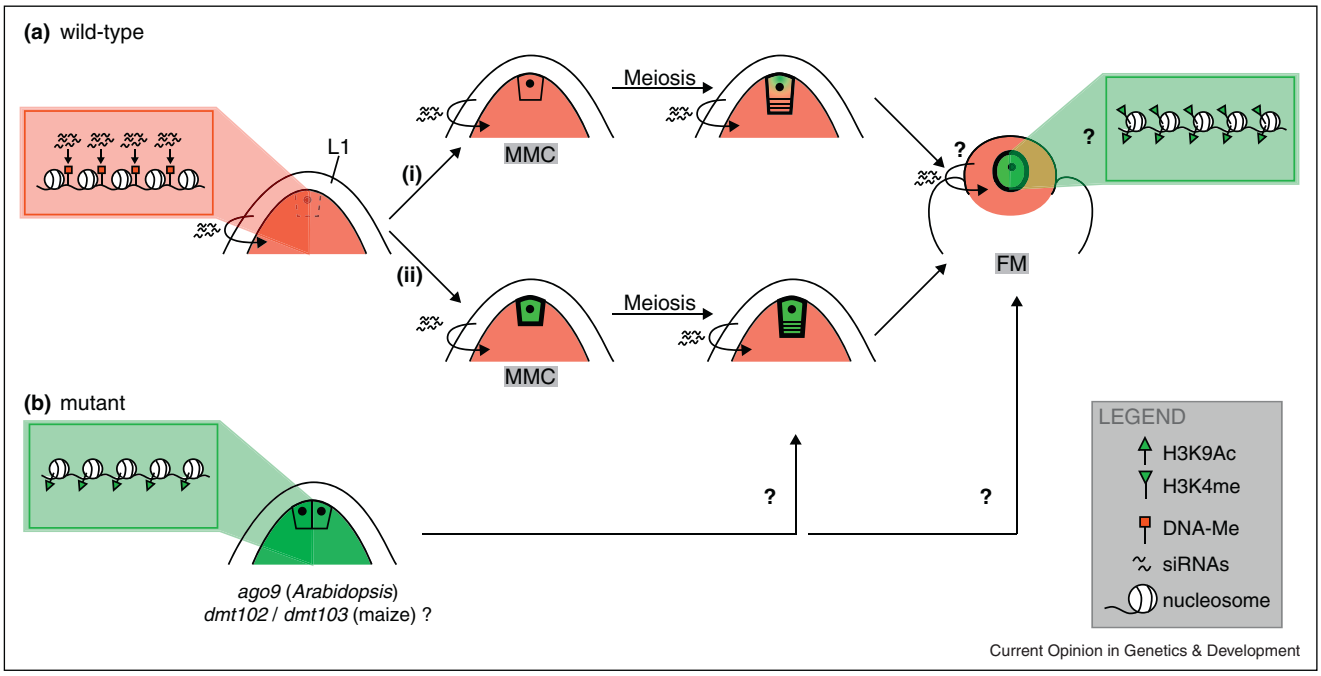
of the ovule [10••,22]. AGO9 belongs to the ARGONAUTE protein family involved in the processing of RNAs into microRNAs (mi-RNAs) and small-interfering RNAs (siRNAs) directing post-transcriptional gene silencing (PTGS) and RNA-dependent DNA methylation (RdDM) [22]. The detection of the AGO9 protein in the epidermal (L1) cell layer specifically suggests that female sporogenic cell fate may be restricted to a single cell via a non-cell autonomous, small RNA-dependent mechanism possibly involving RdDM [10••]. Consistent with this hypothesis, maize nucellar cells express high levels of *Dmt102* transcripts encoding a homologue of the *Arabidopsis* CHROMOMETHYLTRANSFERASE3 (CMT3) [12••], which is responsible for DNA methylation at non-CG sites and acting downstream of siRNA targeting [23]. At the same time and in contrast to adjacent somatic cells, their chromatin is strongly depleted in H3K9Ac [12••], a transcriptionally permissive mark antagonistic to H3K9me2 and DNA methylation, at least in *Arabidopsis* [24]. Down-regulation of *Dmt102*, while accompanied by hyperacetylation of H3K9 in subepidermal cells, was not sufficient to produce supernumerary MMCs as observable at the cytological level, suggesting additional components inhibiting MMC formation [12••]. However, the combined down-regulation of *Dmt102* and *Dmt103*, a close homologue of the *de novo* DNA methyltransferase DOMAIN REARRANGED METHYLTRANSFERASE2 (DRM2), resulted in the formation of ectopic embryo sacs. Some of these may have arisen from supernumerary MMCs [12••], although this awaits confirmation. *Dmt103* is specifically expressed in the epidermal cells of the nucellus [12••]; whether its function is mechanistically connected to an AGO9-related pathway in maize remains to be investigated. This study also revealed the involvement of non-CG methylation (via *Dmt102* and *Dmt103*) rather than CG methylation [12••]. It will be of particular interest to identify the genomic loci affected by siRNA-mediated silencing and non-CG methylation in target cells during sporogenesis.

Profiling at later stages of ovule development revealed that AGO9 preferentially associates with 24nt siRNA targeting transposable elements (TEs), which they regulate in the mature embryo sac, but also with some 20–21nt mi-RNAs species [10••,25]; whether the same small RNAs also act during sporogenesis is yet unknown. To accommodate the observed chromatin patterns in the maize nucellus, one might expect the involvement of small RNAs in targeting — or at least influencing — the silencing in genic regions (not only TEs) to repress sporogenesis and possibly gametogenesis in these cells. The emerging parallel roles of non-coding RNAs controlling germ cell development in animals and plants are of outstanding interest [3•,26]. Whether a similar mechanism is in place during microspore mother cell development should be addressed in the near future. The observation that anther tapetal cells — surrounding the male SMCs and their meiotic daughter cells — produce trans-acting siRNAs targeting the male gametophyte makes this a plausible scenario [27].

Strikingly, the global transcriptionally repressive landscape in the nucellus is reminiscent of the transcriptional quiescence established in the animal germ line. There, it is required to repress the somatic fate but also to reset and establish novel epigenetic profiles in the gametes compatible with totipotency in the zygote [12••,28]. The question is thus which fate is repressed by this apparent transcriptional quiescence in plants? One possibility is the supression of the gametophytic fate (Figure 1, panel (i)). Consistent with this model, the lack of epigenetic repression in the nucellar cells of *Arabidopdis* or maize mutants promotes gametophytic development with the differentiation of additional, albeit abnormal or incomplete, embryo sacs [10••,12••]. Possibly meiosis was bypassed, although more detailed investigations are necessary. If confirmed, this would be reminiscent of apospory, which generates unreduced gametes in some plant species that reproduce through apomixis, the asexual reproduction through seeds [29]. An alternative model invokes the temporal and spatial regulation of sporogenic fate repression (Figure 1, panel (ii)). There, global epigenetic transcriptional repression may first be established in the entire nucellus; later, the selected MMC may escape this repression while it is retained in neighboring cells. In this scenario, siRNA-mediated repression might become less effective in the MMC, possibly as a result of callose deposition in the cell wall. This may constrain intercellular cytoplasmic connections (plasmodesmata) and, thus, transport of siRNAs [30,31]. Consequently, the MMC could recover a transcriptionally permissive state allowing meiosis to take place. Consistent with this hypothesis, meiotic progression — but not entry — is critically dependant on a complex set of histone modifications (reviewed in [3•,32]), including transcriptionally permissive marks established by histone lysine methyltransferases of the SET Domain Group (SDG) [33]. However, the nature of the trigger promoting the mitosis-to-meiosis switch remains unknown. Nonetheless, the prediction that siRNAs were to play an important role in the regulation of meiosis [34] was recently supported by the identification of *MEL1* in rice. *MEL1* encodes an ARGONAUTE protein specifically expressed in SMCs and its loss-of-function causes an arrest at leptotene, the earliest stage where meiosis is distinguished from mitosis [20•]. While the MEL1 clade is closest to the Arabidopsis AGO1-containing clade, its closest homolog in *Arabidopsis*, At2g27880 [20•] was recently annotated as AGO5. The identification of MEL1 targets and their function promises exciting findings with regard to the epigenetic processes controlling the meiotic switch.

Altogether, it becomes clear that sporogenesis is epigenetically controlled both during SMC selection and

Figure 1



Large-scale epigenetic reprogramming during megasporogenesis. **(a)** In ovule primordia, nucellar cells (except for the L1 layer) display a transcriptionally repressive landscape associated with genome-wide histone hypoacetylation, which seems to rely on the RdDM pathway involving AGO9-associated small RNAs originating from L1 cells as well as CMT3-like and DMR2-like activities in nucellar cells. To accommodate mutant phenotypes and expression patterns of epigenetic components as described in the text, two models are proposed that involve dynamic changes of epigenetic landscapes during megasporogenesis. (I) Transcriptional repression inhibits gametophyte development in nucellar cells and the MMC until meiosis is completed. During meiosis permissive histone methylation marks are established that are required for meiotic progression (red-green background). The functional megaspore (FM) has a transcriptionally permissive landscape promoting gametophyte development. (II) Transcriptional repression suppresses the sporogenic fate in nucellar cells. The selected MMC escapes this repression, potentially following isolation from mobile small RNAs. A transcriptionnally permissive landscape is required for completion of meiosis and the promotion of gametophytic development. **(b)** In *Arabidopsis* or maize mutants lacking the function of one epigenetic component mediating transcriptional repression, nucellar cells have a transcriptionally permissive landscape leading to the selection of multiple MMCs and the differentiation of a gametophyte, although abnormal or incomplete.

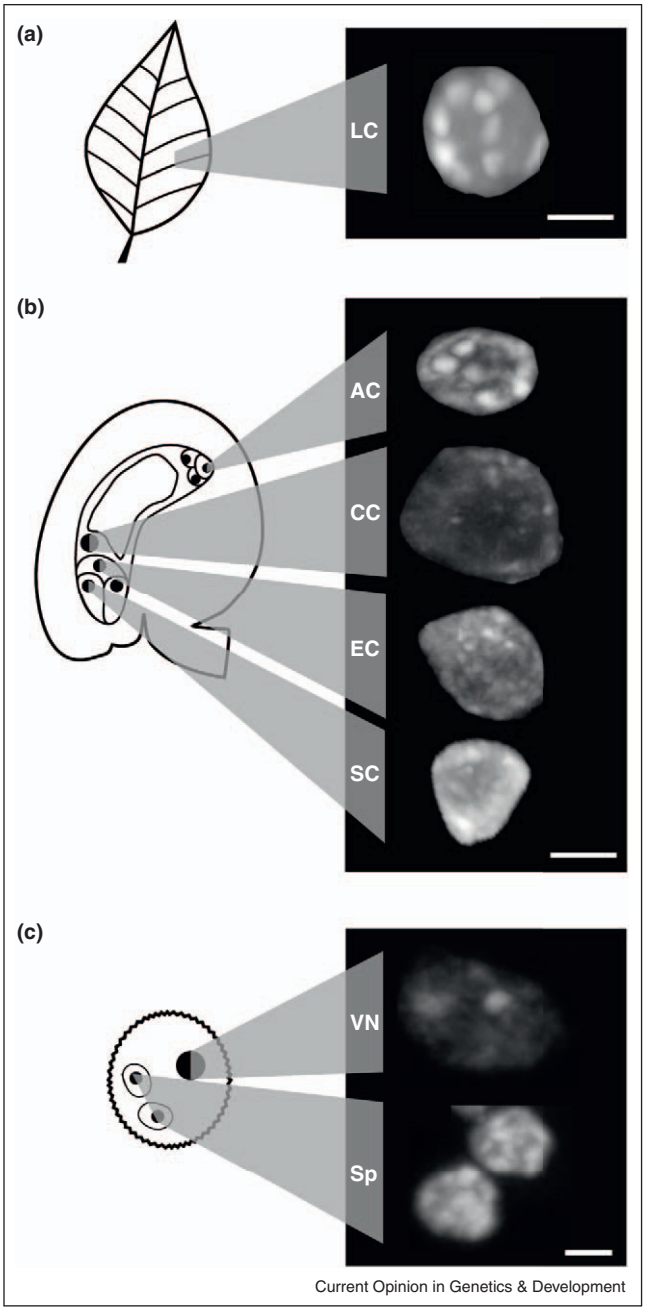
during meiotic progression. Transcriptional quiescence mediated by mobile siRNAs in particular seems instrumental in controlling the sporogenic and possibly the gametophytic fate. Future investigations should aim at revealing the targets of this transcriptional repression, possibly by profiling small RNAs specifically in nucellar cells and MMCs (e.g. using laser-assisted microdissection [35]), but also at elucidating the precise epigenetic chromatin landscape (using immunostaining for different informative histone marks) during sporogenesis.

Epigenetic patterns and reprogramming during gametophyte development

In plants, the meiotic products (spores) enter mitosis and produce a multicelllar, haploid gametophyte, which differentiates gametic as well as accessory cells. Despite the stark reduction and thus simplification of the gametophytes in flowering plants, little is known about the specification of the distinct cell fates [36]. A distinct organization of the chromatin among gametophytic nuclei has been documented early ([1] and Figure 2) but it is only recently that inferences are drawn in terms of epigenetic processes. Several studies now suggest large-scale chromatin modifications associated with the epigenetic differentiation of gametophytic cells.

On the male side, the first asymmetric mitosis generates a small generative cell and a large vegetative cell (Figure 3). The generative cell further divides to produce two sperm cells. Several studies, although fragmented across different species, clearly show that the chromatin composition drastically differs between the sperm and vegetative nucleus. The sperm chromatin, in contrast to the chromatin of the vegetative nucleus, contains gamete-specific core nucleosome variants (gH2A, H2B, gH3) including a gamete-specific H3.3 variant [37–40] (Figure 3). The compact appearance of sperm chromatin (Figure 2) is often interpreted as a transcriptionally inactive state. But, in fact, immunostaining for chromatin modifications reveal a bivalent status: transcriptionally permissive H3K4me3 and H3K36me2/3 marks [41] are below detection level, while euchromatic regions are enriched in

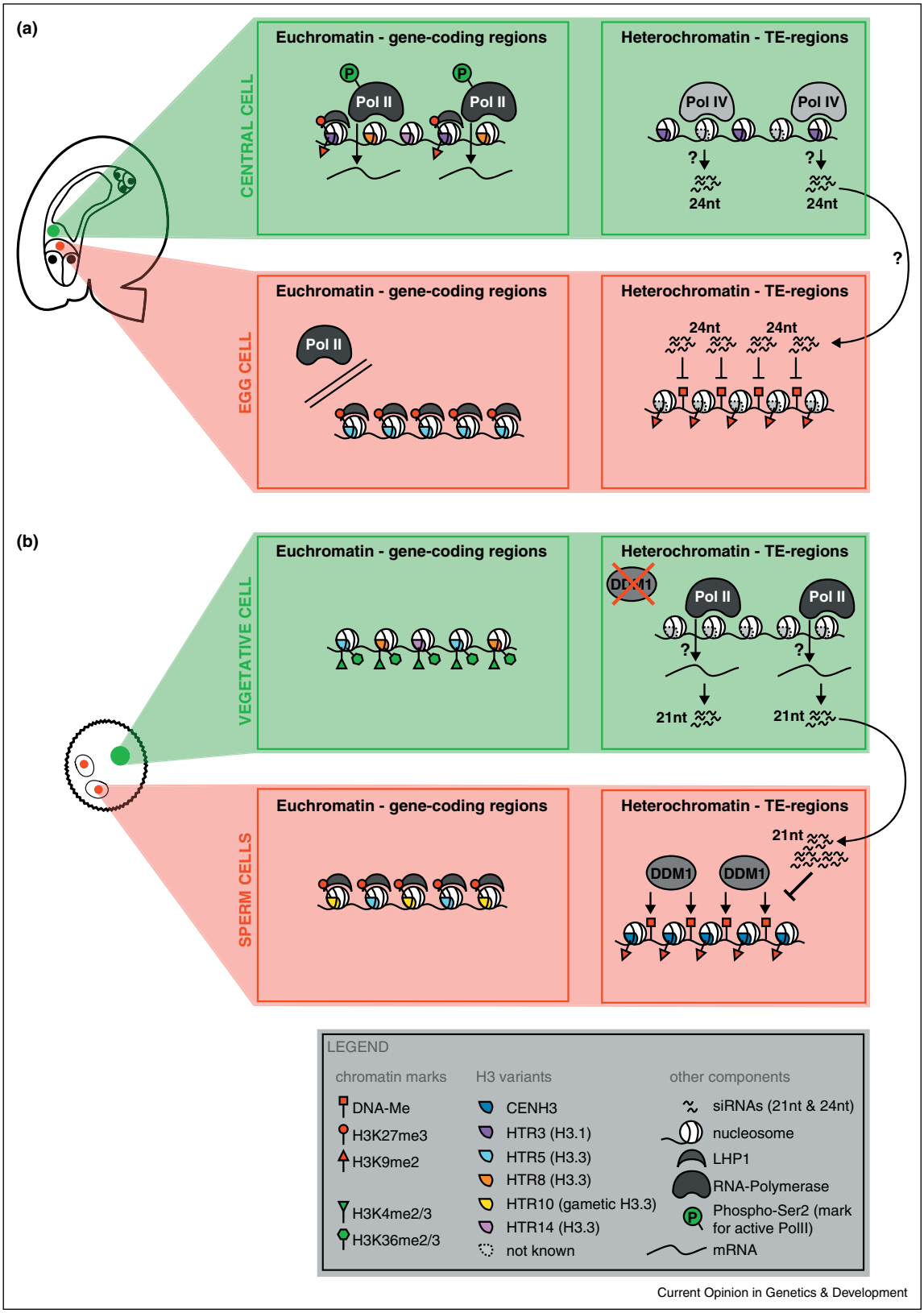
Figure 2



The male and female gametophyte differentiate gametic and accessory cell types with distinct chromatin organization. A distinct organization of the nuclear chromatin in gametophytic nuclei is revealed by whole-mount, non-denaturing DNA staining [77], suggesting that large-scale epigenetic modifications differentiate the mitotic sister cells of the gametophyte (see also Figure 3 and text). **(a)** In sporophytic nuclei, heterochromatin is microscopically visible as well defined, brightly stained foci. **(b)** In the female gametophyte, the gametic nuclei show small, dispersed heterochromatin foci, a feature that is more pronounced in the central cell than the egg cell nucleus. By contrast nuclei of the accessory cells (antipodals and synergids) show higher heterochromatin condensation, with a chromatin organization in antipodal nuclei close to that of a sporophytic cell. **(c)** In the male gametophyte, the vegetative nucleus has largely decondensed chromatin with no apparent heterochromatin organization albeit for a single focus; this organization contrasts with the highly condensed sperm chromatin as inferred by the nuclear size and the discernable heterochromatic foci. This illustrative figure relates to the model presented in Figure 3. All pictures are maximum projections from 3-dimensional reconstructions. Pictures in panel (b) are from C.B. and were reproduced with permission from [78]. Scale bar is 2 μm. AC, antipodal cell; CC, central cell; EC, egg cell; SC, synergid cell; VN, vegetative nucleus; Sp, sperm cells.

LIKE HETEROCHROMATIN PROTEIN1 (LHP1), a protein associated with the transcriptionally repressive H3K27me3 mark [42•,43]. At the same time, hyperacetylation on H3 and H4 residues, together with low levels of DNA methylation [42•,44], suggest that the sperm chromatin retains — to a certain degree — a transcriptionally permissive environment. In stark contrast, the vegetative chromatin appears less condensed, is composed of distinct H3 isoforms, and harbors transcriptionally permissive histone modifications [37,41,42•] (Figures 2 and 3). It also differs in heterochromatin organization and centromeres identity with, particularly, the absence of the centromere-specific CENH3 variant [45] and the heterochromatin-specific somatic H3.1 variants [46•]. This corroborates the absence of the SWI/SNF chromatin remodeler DECREASE DNA METHYLATION1 (DDM1) [9••], normally required for centromeric heterochromatin organization [47]. Thus, clearly, the gametic and accessory chromatin displays an opposite chromatin organization, likely in relation to the resetting of epigenetic marks in the germ line as proposed earlier [3•], but also to the distinct transcriptional programs [48]. The key question remains, however, as to when these differences are established and whether they are cause or consequence of gametic versus accessory cell fate differentiation. The observation that sperm-specific H3 variants are already present in the generative cell [49] suggests that chromatin differentiation is coupled, at least in part, with the first mitosis. Generative and vegetative nuclei clearly show distinct transcriptional patterns supporting their distinct function [48,50]. Misexpression of specific chromatin constituents or modifiers in one mitotic product but not the other may help to understand their influence on the transcriptional programs and downstream differentiation events. Finally, the specific chromatin organization in the vegetative nucleus seems to have a role in the regulation of TEs in the germ line. This was inferred from the observation that the loss of heterochromatin organization results in massive derepression of TEs and the production of 21nt siRNAs, in a DDM1-dependent fashion [9••]. The siRNAs are thought to be transported into the sperm cells where they likely reinforce silencing of TEs. Hence, the vegetative cell may function as a safeguard of genome integrity in the sperm cells in a manner reminiscent of the role of piwiRNAs, 24–31 single-stranded non-coding RNAs principally targeting TEs in the animal germ line [9••,51]. The discovery of a vast and complex population of additional

Figure 3



Distinct epigenetic landscapes regions are established in gametic and accessory nuclei in the gametophytes. This figures combines the findings from several studies on different species but is representative of the situation in *Arabidopsis*. The repertoire of epigenetic modifications shown is not

small RNAs in the male gametophyte, as well as the genetic components producing and targeting them, predicts additional roles in regulating male gametophyte development and repressing somatic functions [52–54].

On the female side, the situation may appear a bit more complex, and certainly less well studied. More complex, because there are three syncytial mitoses followed by nuclear migration and cellularization (Figure 3 represents the most frequent, *Polygonum*-type female gametophyte [1]). Less is known partly because of the difficulty in investigating these cells, which are deeply embedded in maternal tissues. The embryo sac is highly polarized and cellularization establishes distinct cell types (Figure 3). The egg apparatus constituted of the egg cell and two flanking synergids is formed at the micropylar pole, antipodal cells — of yet unknown function — are formed at the chalazal pole, while the central cell in the middle contains two polar nuclei. These migrate towards the center of the embryo sac before cellularization and fuse before or at fertilization depending on the species [7]. These cell types are distinct not only at the cellular, but also at the level of microscopic chromatin organization. Typically, the central and egg cells show a low compaction level together with dispersed heterochromatin foci compared to antipodal and synergid cells (Figure 2, [46•] and C.B., unpublished), suggesting that the gametic versus accessory cell fate may be accompanied by a distinct epigenetic status of the gametophytic cells. It is not yet known whether these distinct features are established before cellularization, in the syncytium, or later. The epigenetic status of these nuclei, defined at the level of the transcriptional pattern, chromatin modifications, and nucleosome constitution, still remains to be analyzed in detail, especially on a temporal scale along with mitotic progression, polarity establishment, and cellularization. Chromatin changes affecting nucleosome constitution and biochemical modifications appear to occur dynamically following cellularization [11•,46•] and contribute to establish a distinct transcriptional status in egg and central cell (see below). Conversely, DNA methylation patterns of gametophytic nuclei may regulate cell identity, cell number and embryo sac polarity as shown in the *dmt102* and *dmt103* maize mutants [12••]. To date, no such phenotypes have been documented in the corresponding *Arabidopsis* mutants, which are fully fertile [55,56] suggesting species-dependent epigenetic control during gametophyte development.

Epigenetic dimorphism between the female gametes

Double fertilization in flowering plants poses the intriguing problem to (epigenetically) distinguish the developmental fates of the fertilization products. In most species each pair of male or female gametes arise from a single male or female spore, respectively. In most species the two sperm are cytologically indistinguishable and can fertilize either the central or egg cell [57,58]. Thus, it was proposed that the female gametes must have different epigenetic states that contribute to the distinct post-fertilization fates [21,8]. But only recently detailed chromatin investigations revealed a stark epigenetic dimorphism between the two female gametes, in both *Arabidopsis* and maize [11•,12••,46•]. The egg cell displays a global transcriptionally quiescent chromatin state with undetectable levels of active RNA Polymerase II (PolII) concomitant with enrichment in LHP1. By contrast the central cell chromatin displays a transcriptionally active and permissive chromatin organization [10••,11•], Figure 3). At the same time, microscopically visible heterochromatin regions, mostly composed of TEs [59], are well defined and enriched in H3K9me2 in the egg as compared to the central cell where heterochromatic foci are less pronounced (Figure 2 and [46•]). In *Arabidopsis*, this dimorphism requires the activity of CMT3 in the egg cell, and DEMETER-LIKE enzymes, catalyzing the removal of methylcytosine residues, in the central cell [11•,42•]. Importantly, the distinct global epigenetic

(Figure 3 Legend Continued) exhaustive and only represents studied marks. **(a)** In the female gametophyte the two female gametes show an epigenetic dimorphism. In comparison to the egg cell, the central cell is abundant for the active form of RNA Pol II, features a transcriptionally permissive environment in euchromatic regions with quantitatively little repressive marks, as inferred from LHP1 levels (LHP1 is a H3K27me3-associated protein [43]). By contrast the egg's euchromatin displays a transcriptionally quiescent state with opposite active Pol II and LHP1 abundance (note that the inactive form of Pol II is present in the egg nucleus). Distinct variants of H3.3 histones are found in the egg and central cell but their role in transcriptional control is not known. At heterochromatin regions, the scenario is hypothetical and based on observations made at later stage of seed development (see text), mirroring a similar scenario in the male gametophyte **(b)**. The central cell is proposed to produce small RNAs potentially targeting TEs in the egg cell. The lack of heterochromatin at TE regions linked to a lack of centromeric identity (dispersed heterochromatin foci [Figure 2] and absence of detectable CENH3 (see text) may favor reactivation of TEs. This hypothetically involves RNA Pol IV to produce 24nt-siRNAs as they were detected at late stages of endosperm development. Potentially, these siRNA are transported into the egg cell to ensure TE silencing via RdDM as was proposed for the embryo at later stages, and they might be linked to high levels of H3K9me2 as observed in the egg cell. (b) In the male gametophyte, epigenetic dimorphism is not found between the gametes but rather between the sperm cells and the vegetative cell. Euchromatin regions show histone modifications linked with a transcriptionally permissive landscape in the vegetative nucleus, in contrast to the euchromatin in sperm cells. At the nuclesome level, they both show a common histone H3.3 variant (HTR5) but also distinct variants, including a gamete-specific variant (HTR10) in the sperm. Whether these variants influence the epigenetic state is not known. At heterochromatin regions, the absence of the SWI/SNF chromatin remodeler DDM1 and the loss of centromeric heterochromatin organization is correlated with derepression of TEs, transcription by RNA Pol IV and production of 21nt small RNAs accumulating in the sperm cells. These siRNAs are thought to reinforce TE silencing in the gametes, possibly via RdDM, to ensure genome integrity. This correlates with enrichment of H3K9me2 at heterochromatic foci, although contradictory immunostaining results were reported [41,42•].

states affecting chromatin modifications and transcription are established soon after cellularization and are inherited post-fertilization [11•]. This dimorphism also correlates with a genome-wide hypomethylation found later during seed development in the endosperm as compared to the embryo, in both *Arabidopsis* and maize [60–62]. Epigenetic dimorphism between the female gametes is further found at the level of nucleosome composition with specific sets of H3.3 isoforms in each cell, as inferred from tagged histone reporters ([46•], Figure 3). Particularly, the egg cell displays only one H3.3 isoform (HTR5) also found in the sperm cells, while the central cells displays two isoforms (HTR8 and HTR14). Whether HTR8 and HTR14 epigenetically distinguish two functional chromatin compartments would be interesting, yet challenging to determine. Interestingly, this nucleosome dimorphism between the egg and central cell is ephemeral and seems remodeled after fertilization, suggesting that it was established solely for the purpose of distinguishing the egg versus central cell chromatin during gametophyte development. This also implies that maternal epigenetic marks associated with these H3 variants are not transmitted to the zygote [46•]. But the possibility remains that some loci escape this remodeling, a situation that would not be detected at the microscopic scale with reporter histones such as used in this study [46•], and leading to the transmission of a maternal epigenetic mark (imprint). Similarly, for some loci, especially those showing parent-of-origin-specific expression following fertilization, the epigenetic states established in the female gametes may be dynamically modified following fertilization [63,64]. For instance, DNA methylation patterns established on the maize *ZmFIE1 and Mee1* loci in the central and egg cells, respectively, are not maintained in the fertilization products [63,64].

But what could explain this epigenetic dimorphism? The distinct transcriptional states support the different requirements of the fertilization products on maternally inherited information: while the endosperm strictly requires *de novo* transcription, the quiescent zygote can undergo 4–5 mitoses in the absence of active PolII [11•]. Initial quiescence in the plant zygote is reminiscent of the situation in animals, where it is a strict requirement for genome reprogramming and the acquisition of totipotency [28]. However, releasing the quiescence in the egg cell is not lethal for the embryo [42•]. Instead, transient defects in patterning are observed [42•], which could be explained by the relative robustness of the plant genome to epigenetic perturbations, possibly mediated by alternative epigenetic pathways [42•,65]. Thus, the relative quiescence in the egg, possibly linked to the selective reprogramming of maternal epigenetic information, might fulfill additional roles. These may be revealed in natural, ecological contexts under selective pressures, by contrast to laboratory conditions where history of the maternal parent does not obviously influence the offspring (at least not detectably). Finally, epigenetic dimorphisms at the level of TE-containing heterochromatic regions may support a similar role to that described in the male gametophyte. In such a model, the central cell is proposed to contribute to protecting genome integrity in the egg by producing and transporting siRNAs reinforcing TE silencing [66,67,68]. This model arose from combined observations made at later stage during seed development, in the endosperm at 7–9 days after pollination. A complex set of maternal small RNAs, including 24nt (and not 21nt as in the pollen [9••]) siRNAs mostly targeting TEs, was found [69•] concomitant with an extensive demethylation of these elements [60]. While very attractive, this model awaits demonstration especially for the presence and mobility of such small RNAs in the central cell. TEs are derepressed in *ago9* but also *cmt3* mutant gametophytes [10••,42•], comforting the hypothesis that small RNAs may preserve genome integrity in the egg cell. However, the expression pattern of AGO9 also suggests that the maternal sporophyte may be involved in this process [10••,25].

Concluding remarks

The past decade of research revealed exciting findings with regard to epigenetic mechanisms controlling developmental processes specific to flowering plants: the determination of sporogenic fate late during development, the differentiation of gametes within multicellular gametophytes, and the distinction of the two female gametes involved in double fertilization. Particularly, there seem to be two levels of epigenetic regulation, both acting at large-scale: one regulatory level targets euchromatic regions where a transcriptionally repressive landscape seems to be established during both sporogenic fate acquisition and egg — and likely sperm — differentiation. This is reminiscent of the quiescence in the animal germ line and early zygote, necessary for the acquisition of totipotency [28]. The involvement of mobile non-coding siRNAs, contributed by the maternal sporophyte or the sporogenic/gametophytic lineage itself, is particularly intriguing and will surely stimulate future research in the field. Furthermore, whether transcriptional quiescence during sporogenesis and gametogenesis plays a similar role in plants and animals remains to be determined. It will be of particular interest to determine whether the epigenetic information related to the maternal history — and thus linked to the environment — is maintained or reset during these phases. On the one hand, the ability to flower acquired following vernalization (winter-mimicking period of cold), and mediated by the epigenetic repression of the flowering control locus FLC, has to be reset in the new diploid generation [70–72]. On the other hand, explaining some maternal effects on the progeny may require the inheritance of maternally acquired epigenetic information [73,74]; although virtually nothing is known about the mechanism. The

second level of epigenetic regulation emerging from recent studies is the control of TEs mostly contained in microscopically visible heterochromatic regions and the centromeres of *Arabidopsis* [59]. In order to preserve genome integrity in the developing embryo, plants seem to sacrifice sister cells of the gametes — the accessory vegetative cell in the pollen and possibly the central cell in the embryo sac — which do not contribute to the next generation. In the current model, which still awaits demonstration, these cells produce mobile siRNAs reinforcing TE silencing in the gametes and possibly in the embryo.

Finally, and on a different note, the recent findings presented in the first section have important implications in the understanding of apomixis, a naturally occurring mode of asexual reproduction through seeds, involving the production of additional unreduced gametophytes in some plant species. Particularly, these findings are consistent with the proposed epigenetic basis for the deregulation of sporogenesis and/or gametogenesis during apomixis [29,75,76].

Acknowledgements

We thank D. Grimanelli for stimulating discussions on topics included in this review. Our work on the epigenetic regulation of plant reproduction is supported by the University of Zürich, a fellowship of the Roche Research Foundation (to MTR), the EPIGENOME Network of Excellence and the European Research Council (to UG), and grants of the Swiss National Science Foundation (to UG and CB).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Maheshwari P: *An Introduction to the Embryology of Angiosperms*. New York/Toronto/London: McGraw-Hill Book Company; 1950.

2. Walbot V, Evans MM: **Unique features of the plant life cycle and their consequences**. *Nat Rev Genet* 2003, **4**:369-379.

3. • Dickinson HG, Grant-Downton R: **Bridging the generation gap: flowering plant gametophytes and animal germlines reveal unexpected similarities**. *Biol Rev Camb Philos Soc* 2009, **84**:589-615.
This work exhaustively reviews cytological, molecular, and epigenetic aspects of male and female gametophyte development in flowering plants, while drawing interesting comparisons to animal systems.

4. Banks JA: **Gametophyte development in ferns**. *Annu Rev Plant Physiol Plant Mol Biol* 1999, **50**:163-186.

5. Rudall PJ: **How many nuclei make an embryo sac in flowering plants?** *Bioessays* 2006, **28**:1067-1071.

6. Rudall PJ, Bateman RM: **Developmental bases for key innovations in the seed–plant microgametophyte**. *Trends Plant Sci* 2007, **12**:317-326.

7. Dumas C, Rogowsky P: **Fertilization and early seed formation**. *C R Biol* 2008, **331**:715-725.

8. Messing J, Grossniklaus U: **Genomic imprinting in plants**. *Results Probl Cell Differ* 1999, **25**:23-40.

9. •• Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA: **Epigenetic reprogramming and small RNA silencing of transposable elements in pollen**. *Cell* 2009, **136**:461-472.
This is the first report showing a role for non-coding siRNAs in the regulation of transposable elements (TEs) in the male germ line, that is in *Arabidopsis* pollen. The data support a model where 21nt siRNAs targeting TEs are generated in the vegetative cell and mobilized to the sperm cells where they would reinforce TE silencing, and hence preserve genome integrity.

10. •• Olmedo-Monfil V, Duran-Figueroa N, Arteaga-Vazquez M, Demesa-Arevalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP: **Control of female gamete formation by a small RNA pathway in Arabidopsis**. *Nature* 2009, **464**:628-632.
The authors identified that a loss of *ARGONAUTE9* function in *Arabidopsis* deregulates sporogenic fate control with the production of multiple MMCs and ectopic embryo sacs. Furthermore, transposable elements are derepressed in *ago9* embryo sacs, highlighting an additional role in transposable element regulation in the gametes. Expression of *AGO9* in the epidermal cells suggests a non-cell autonomous pathway in both processes.

11. • Pillot M, Baroux C, Vazquez MA, Autran D, Leblanc O, Vielle-Calzada JP, Grossniklaus U, Grimanelli D: **Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in Arabidopsis**. *Plant Cell* 2010, **22**:307-320.
The female gametes involved in double fertilization harbor a stark epigenetic dimorphism at the chromatin and transcriptional level, with the egg showing a relative quiescence reminiscent of the situation in animals. This dimorphism is genetically controlled in the egg by CMT3 directing non-CG methylation and in the central cell by DEMETER-LIKE enzymes mediating removal of methylcytosine residues.

12. •• Garcia-Aguilar M, Michaud C, Leblanc O, Grimanelli D: **Inactivation of a DNA methylation pathway in maize reproductive organs results in apomixis-like phenotypes**. *Plant Cell* 2010, **22**:307–320.
This study identifies four loci involved in gene silencing via DNA methylation that are deregulated between sexual and apomictic maize lines. It further demonstrates a role for two of them, encoding CMT3 and DRM2 homologues, during sporogenesis and gametogenesis and in the establishment of transcriptionnally repressive patterns during these processes.

13. Stahl Y, Simon R: **Plant primary meristems: shared functions and regulatory mechanisms**. *Curr Opin Plant Biol* 2010, **13**:53-58.

14. Rédei GP: **Analysis of the diploid germline of plants by mutational techniques**. *Can J Genet Cytol* 1974, **16**:473-476.

15. Sheridan WF, Golubeva EA, Abrhamova LI, Golubovskaya IN: **The mac1 mutation alters the developmental fate of the hypodermal cells and their cellular progeny in the maize anther**. *Genetics* 1999, **153**:933-941.

16. Nonomura K, Miyoshi K, Eiguchi M, Suzuki T, Miyao A, Hirochika H, Kurata N: **The MSP1 gene is necessary to restrict the number of cells entering into male and female sporogenesis and to initiate anther wall formation in rice**. *Plant Cell* 2003, **15**:1728-1739.

17. Terpstra I, Heidstra R: **Stem cells: the root of all cells**. *Semin Cell Dev Biol* 2009, **20**:1089-1096.

18. Sheridan WF, Avalkina NA, Shamrov II, Batygina TB, Golubovskaya IN: **The mac1 gene: controlling the commitment to the meiotic pathway in maize**. *Genetics* 1996, **142**:1009-1020.

19. Zhao DZ, Wang GF, Speal B, Ma H: **The excess microsporocytes1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the Arabidopsis anther**. *Genes Dev* 2002, **16**:2021-2031.

20. • Nonomura K, Morohoshi A, Nakano M, Eiguchi M, Miyao A, Hirochika H, Kurata N: **A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice**. *Plant Cell* 2007, **19**:2583-2594.
The authors describe the rice mutant *mel1,* affected in an *ARGONAUTE* family member closely related to *AGO1* and *AGO5*, causing early meiotic arrest during sporogenesis. The arrest at leptotene in early prophase I suggests that *MEL1* may be involved in the mitotic-meiotic switch in rice, a key transition during sporogenesis. It further suggests a role for non-coding siRNA-mediated regulation in sporogenesis.

21. Grossniklaus U, Schneitz K: **The molecular and genetic basis of ovule and megagametophyte development**. *Semin Cell Dev Biol* 1998, **9**:227-238.

22. Vaucheret H: **Plant ARGONAUTES**. *Trends Plant Sci* 2008, **13**:350-358.

23. Chan SW: **Inputs and outputs for chromatin-targeted RNAi**. *Trends Plant Sci* 2008, **13**:383-389.

24. Zhou J, Wang X, He K, Charron JB, Elling AA, Deng XW: **Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in Arabidopsis reveals correlation between multiple histone marks and gene expression**. *Plant Mol Biol* 2010, **72**:585-595.

25. Duran-Figueroa N, Vielle-Calzada JP: **ARGONAUTE9-dependent silencing of transposable elements in pericentromeric regions of Arabidopsis**. *Plant Signal Behav* 2010:5.

26. Tang F: **Small RNAs in mammalian germline: tiny for immortal**. *Differentiation* 2010, **79**:141-146.

27. Tarutani Y, Shiba H, Iwano M, Kakizaki T, Suzuki G, Watanabe M, Isogai A, Takayama S: **Trans-acting small RNA determines dominance relationships in Brassica self-incompatibility**. *Nature* 2010, **466**:983-986.

28. Seydoux G, Braun RE: **Pathway to totipotency: lessons from germ cells**. *Cell* 2006, **127**:891-904.

29. Koltunow AM, Grossniklaus U: **Apomixis: a developmental perspective**. *Annu Rev Plant Biol* 2003, **54**:547-574.

30. Rodkiewicz B: **Callose in cell walls during megasporogenesis in Angiosperms**. *Planta* 1970, **93**:39-47.

31. Xu XM, Jackson D: **Lights at the end of the tunnel: new views of plasmodesmal structure and function**. *Curr Opin Plant Biol* 2010, **13**:684-692.

32. Ronceret A, Sheehan MJ, Pawlowski WP: **Chromosome dynamics in meiosis**. In *Cell Division Control in Plants*, vol 9. Edited by Zong DPSV. 2010:103-124. Plant Cell Monogr.

33. Berr A, McCallum EJ, Menard R, Meyer D, Fuchs J, Dong A, Shen WH: **Arabidopsis SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development**. *Plant Cell* 2010, **22**:3232-3248.

34. Hamant O, Ma H, Cande WZ: **Genetics of meiotic prophase I in plants**. *Annu Rev Plant Biol* 2006, **57**:267-302.

35. Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenfuhrer J, von Mering C, Grossniklaus U: **Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes**. *Curr Biol* 2010, **20**:506-512.

36. Ma H, Sundaresan V: **Development of flowering plant gametophytes**. *Curr Top Dev Biol* 2010, **91**:379-412.

37. Okada T, Endo M, Singh MB, Bhalla PL: **Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3**. *Plant J* 2005, **44**:557-568.

38. Okada T, Singh MB, Bhalla PL: **Histone H3 variants in male gametic cells of lily and H3 methylation in mature pollen**. *Plant Mol Biol* 2006, **62**:503-512.

39. Ueda K, Kinoshita Y, Xu ZJ, Ide N, Ono M, Akahori Y, Tanaka I, Inoue M: **Unusual core histones specifically expressed in male gametic cells of *Lilium longiflorum***. *Chromosoma* 2000, **108**:491-500.

40. Xu H, Swoboda I, Bhalla PL, Singh MB: **Male gametic cell-specific expression of H2A and H3 histone genes**. *Plant Mol Biol* 1999, **39**:607-614.

41. Cartagena JA, Matsunaga S, Seki M, Kurihara D, Yokoyama M, Shinozaki K, Fujimoto S, Azumi Y, Uchiyama S, Fukui K: **The Arabidopsis SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen**. *Dev Biol* 2008, **315**:355-368.

42. • Pillot M, Autran D, Leblanc O, Grimanelli D: **A role for CHROMOMETHYLASE3 in mediating transposon and euchromatin silencing during egg cell reprogramming in Arabidopsis**. *Plant Signal Behav* 2010:5.

In this work, the authors show that CMT3 is, similar to AGO9, involved in transposable element regulation in the female gametophyte. This connects siRNA-mediated silencing and non-CG methylation in this process.

43. Exner V, Aichinger E, Shu H, Wildhaber T, Alfarano P, Caflisch A, Gruissem W, Kohler C, Hennig L: **The chromodomain of LIKE HETEROCHROMATIN PROTEIN 1 is essential for H3K27me3 binding and function during Arabidopsis development**. *PLoS One* 2009, **4**:e5335.

44. Januseck B, Zluvova J, Vyskot B: **Histone H4 acetylation and DNA methylation dynamics during pollen development**. *Protoplasma* 2005, **211**:116-122.

45. Schoft VK, Chumak N, Mosiolek M, Slusarz L, Komnenovic V, Brownfield L, Twell D, Kakutani T, Tamaru H: **Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin**. *EMBO Rep* 2009, **10**:1015-1021.

46. • Ingouff M, Rademacher S, Holec S, Soljic L, Xin N, Readshaw A, Foo SH, Lahouze B, Sprunck S, Berger F: **Zygotic Resetting of the HISTONE 3 Variant Repertoire Participates in Epigenetic Reprogramming in Arabidopsis**. *Current Biology* 2010, **20**:2137–2143.

This study report the generation of tagged histone H3 variants covering the entire repertoire of expressed isoforms in *Arabidopsis*. This allowed describing specific sets of variants in the male and female gametes. The authors further show a dynamic remodeling of the tagged histones inherited by either parent after fertilization.

47. Soppe WJ, Jasencakova Z, Houben A, Kakutani T, Meister A, Huang MS, Jacobsen SE, Schubert I, Fransz PF: **DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis**. *EMBO J* 2002, **21**:6549-6559.

48. Borg M, Brownfield L, Twell D: **Male gametophyte development: a molecular perspective**. *J Exp Bot* 2009, **60**:1465-1478.

49. Brownfield L, Hafidh S, Borg M, Sidorova A, Mori T, Twell D: **A plant germline-specific integrator of sperm specification and cell cycle progression**. *PLoS Genet* 2009, **5**:e1000430.

50. Okada T, Singh MB, Bhalla PL: **Transcriptome profiling of *Lilium longiflorum* generative cells by cDNA microarray**. *Plant Cell Rep* 2007, **26**:1045-1052.

51. Bourc'his D, Voinnet O: **A small-RNA perspective on gametogenesis, fertilization, and early zygotic development**. *Science* 2010, **330**:617-622.

52. Grant-Downton R, Hafidh S, Twell D, Dickinson HG: **Small RNA pathways are present and functional in the angiosperm male gametophyte**. *Mol Plant* 2009, **2**:500-512.

53. Grant-Downton R, Le Trionnaire G, Schmid R, Rodriguez-Enriquez J, Hafidh S, Mehdi S, Twell D, Dickinson H: **MicroRNA and tasiRNA diversity in mature pollen of *Arabidopsis thaliana***. *BMC Genomics* 2009, **10**:643.

54. Le Trionnaire G, Twell D: **Small RNAs in angiosperm gametophytes: from epigenetics to gamete development**. *Genes Dev* 2010, **24**:1081-1085.

55. Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE: **Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation**. *Science* 2001, **292**:2077-2080.

56. Cao X, Jacobsen SE: **Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing**. *Curr Biol* 2002, **12**:1138-1144.

57. Ingouff M, Sakata T, Li J, Sprunck S, Dresselhaus T, Berger F: **The two male gametes share equal ability to fertilize the egg cell in *Arabidopsis thaliana***. *Curr Biol* 2009, **19**:R19-20.

58. Faure JE, Rusche ML, Thomas A, Keim P, Dumas C, Mogensen HL, Rougier M, Chaboud A: **Double fertilization in maize: the two male gametes from a pollen grain have the ability to fuse with egg cells**. *Plant J* 2003, **33**:1051-1062.

59. Fransz P, Soppe W, Schubert I: **Heterochromatin in interphase nuclei of *Arabidopsis thaliana***. *Chromosome Res* 2003, **11**:227-240.

60. Gehring M, Bubb KL, Henikoff S: **Extensive demethylation of repetitive elements during seed development underlies gene imprinting**. *Science* 2009, **324**:1447-1451.

61. Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D: **Genome-wide demethylation of Arabidopsis endosperm**. *Science* 2009, **324**:1451-1454.

62. Lauria M, Rupe M, Guo M, Kranz E, Pirona R, Viotti A, Lund G: **Extensive maternal DNA hypomethylation in the endosperm of *Zea mays***. *Plant Cell* 2004, **16**:510-522.

63. Gutierrez-Marcos JF, Costa LM, Dal Pra M, Scholten S, Kranz E, Perez P, Dickinson HG: **Epigenetic asymmetry of imprinted genes in plant gametes**. *Nat Genet* 2006, **38**:876-878.

64. Jahnke S, Scholten S: **Epigenetic resetting of a gene imprinted in plant embryos**. *Curr Biol* 2009, **19**:1677-1681.

65. Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J: **Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation**. *Cell* 2007, **130**:851-862.

66. Martienssen RA: **Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of Arabidopsis**. *New Phytol* 2010, **186**:46-53.

67. Springer NM: **Small RNAs: how seeds remember to obey their mother**. *Curr Biol* 2009, **19**:R649-651.

68. Mosher RA, Melnyk CW: **siRNAs and DNA methylation: seedy epigenetics**. *Trends Plant Sci* 2010, **15**:204-210.

69. • Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC: **Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis**. *Nature* 2009, **460**:283-286.
This study reveals the presence of a complex set of maternally derived siRNAs in the endosperm of developed seeds, targeting essentially transposable elements. These siRNAs are dependent on the maternal activity of RNA PolIV.

70. Feng S, Jacobsen SE, Reik W: **Epigenetic reprogramming in plant and animal development**. *Science* 2010, **330**:622-627.

71. Choi J, Hyun Y, Kang MJ, In Yun H, Yun JY, Lister C, Dean C, Amasino RM, Noh B, Noh YS *et al.*: **Resetting and regulation of Flowering Locus C expression during Arabidopsis reproductive development**. *Plant J* 2009, **57**:918-931.

72. Sheldon CC, Hills MJ, Lister C, Dean C, Dennis ES, Peacock WJ: **Resetting of FLOWERING LOCUS C expression after epigenetic repression by vernalization**. *Proc Natl Acad Sci U S A* 2008, **105**:2214-2219.

73. Munir J, Dorn LA, Donohue K, Schmitt J: **The effect of maternal photoperiod on seasonal dormancy in *Arabidopsis thaliana* (Brassicaceae)**. *Am J Bot* 2001, **88**:1240-1249.

74. Galloway LF, Etterson JR: **Transgenerational plasticity is adaptive in the wild**. *Science* 2007, **318**:1134-1136.

75. Grimanelli D, Garcia M, Kaszas E, Perotti E, Leblanc O: **Heterochronic expression of sexual reproductive programs during apomictic development in Tripsacum**. *Genetics* 2003, **165**:1521-1531.

76. Grossniklaus U, Nogler GA, van Dijk PJ: **How to avoid sex: the genetic control of gametophytic apomixis**. *Plant Cell* 2001, **13**:1491-1498.

77. Baroux C, Pecinka A, Fuchs J, Schubert I, Grossniklaus U: **The triploid endosperm genome of Arabidopsis adopts a peculiar, parental-dosage-dependent chromatin organization**. *Plant Cell* 2007, **19**:1782-1794.

78. Kohler C, Gaudin V, Hennig L: **Green chromatin dynamics in Zurich: meeting summary based on the European Workshop on Plant Chromatin 2009 in Zurich, Switzerland**. *Epigenetics* 2010, **5**:80-83.